

# Inhaltsverzeichnis

Vorwort .....	ix
---------------	----

## „Open Innovation“

### Neue Perspektiven im Kontext von Information und Wissen

#### 10. Internationales Symposium für Informationswissenschaft

### Aktuelle Fragen und Experimente des Information Retrieval

*Fabrizio Falchi, Fausto Rabitti, Werner Schweibenz & Jan Simane*

Florentine Coats of Arms on the Web: Experimenting with Retrieval Based on Text or Image Content .....	1
---	---

*Kerstin Bischoff, Ralph Kölle, Thomas Mandl & Christa Womser-Hacker*

Geographische Bedingungen im Information Retrieval: Neue Ansätze in Systementwicklung und Evaluierung.....	15
---	----

### Wissensvermittlung im digitalen Umfeld

*Tobias Galliat*

Personalisierung mittels informationsbewertender Profile.....	27
---	----

*Maria Baumgartner & Christian Schlögl*

Die Tagungsbände des Internationalen Symposiums für Informationswissen- schaft in szientometrischer Analyse .....	43
--	----

*Susanne Mühlbacher*

Meta-Strategies to foster a Meta-Competence? Supporting information literacy at the workplace .....	61
--	----

*Rainer Hammwöhner, Karl-Peter Fuchs, Markus Kattenbeck & Christian Sax*

Qualität der Wikipedia – eine vergleichende Studie.....	77
---	----

### Open Access Publishing

*Dietmar Schiller, Juliane Burghardt, Nadine Fijalkowytch, Veit-E. Jauß &  
Bodo Schindler*

Vom “Public Service” zum “Public Value” – Öffentlich-rechtliche Programmarchive der Zukunft als digitale Wissensspeicher.....	91
--	----

*A.R.D. Prasad & Devika P. Madalli*

OPEN MANTRA for Open Access to Information.....	96
---	----

*Achim Oßwald*

Aufbau und Betrieb digitaler Repositorien in Indien ..... 109

## **Multimedialität und Multimodalität**

*Klaus D. Stiller*

The Modality Principle in Multimedia Learning..... 129

*Rüdiger Heimgärtner*

Cultural Differences in Human Computer Interaction:  
Results from Two Online Surveys ..... 145

*Andreas Ratzka*

A Wizard-of-Oz Setting for Multimodal Interaction ..... 159

## **Information Management in Unternehmen und öffentlichen Einrichtungen**

*Monika Maßun*

From Personal Information Management  
Towards Collaborative Information Management in Enterprises..... 171

*Ragna Seidler-de Alwis*

Competitive Intelligence und Informationsvermittlung in  
deutschen Unternehmen..... 189

*Christine Krätzsch*

Weblogs als Steuerungsinstrument in Hochschulbibliotheken ..... 207

## **Digitalisierung und Open Innovation**

*Christina Bieber*

Von der Digitalisierung der Objekte zur Virtualisierung der Kulturgüter ..... 221

*Theodora Stathoulia*

Sceptical Doubts Concerning Knowledge Capture within the  
Open Innovation Initiative..... 237

*Jakob Voß*

Tagging, Folksonomy & Co. – Renaissance of Manual Indexing? ..... 243

## **E-Learning**

*Joachim Griesbaum, Wolfgang Semar, Tao Jiang & Rainer Kuhlen*

Feedback zur Beförderung netzbasierter Wissenskommunikation in K3 ..... 255

*Thomas Sporer, Ulrich Fahrner & Lutz Mauermann*

Gestaltung von digitalen Lehr-Lern-Szenarien durch offene  
Innovationsstrategien an Hochschulen..... 271

## **Semantic Web / Social Software**

*Jürgen Reischer*

Extracting Informative Content Units in Text Documents ..... 285

*Christopher Lueg*

Gathering Information in Online Communities..... 303

## **Poster und Systemdemonstrationen**

*Wiebke Oeltjen*

Content Management mit MyCoRe Dokumenten- und  
Publikationsserver und Archivanwendungen mit der  
Open Source Software MyCoRe ..... 313

*Wolfgang Semar, Joachim Griesbaum, Tao Jiang & Rainer Kuhlen*

K3 – Ein kollaboratives Wissensmanagementsystem zur Beförderung von  
Kommunikations- und Informationskompetenz..... 317

*Anton Köstlbacher, Christoph Hiemke, Ekkehard Haen, Gabriel Eckermann,  
Matthias Dobmeier & Rainer Hammwöhner*

PsiaOnline – Fachdatenbank für Arzneimittelwechselwirkungen in der  
psychiatrischen Pharmakotherapie ..... 321

*Paul Landwich, Norbert Fuhr & Matthias Hemmje*

Ansatz zu einem konzeptionellen Modell für interaktive  
Information Retrieval-Systeme mit Unterstützung von  
Informationsvisualisierung ..... 327

*Sonja Öttl, Sonja Hierl, Bernard Bekavac & Josef Herget*

Die Analyse heterogener Unternehmensdatenbestände als Basis für die  
Visualisierung von Relationen in Suchergebnismengen..... 333

*Christian Wolff*

Informationskompetenz als Studienfach?!  
Eine curriculare Kooperation von Universitätsbibliothek und  
Fachdisziplinen ..... 343

*Elke Bernsee & Christoph Nitrowski*

Wissenschaftsportal b2i – Bibliotheks-, Buch- und  
Informationswissenschaften: Aktueller Stand des Projektes und Ausblick ..... 351

## Poster informationswissenschaftlicher Doktoranden

*Wolfgang Granigg*

Semantik in Informationssystemen..... 355

*Christian Rengstl*

Genome-wide Clustering of Genes..... 361

*Nadine Höchstötter*

Untersuchung des Online Suchverhaltens im Web Interaktion  
von Internetnutzern mit Suchmaschinen ..... 367

*Daniel Jobst*

Prozess- und Ereignisorientierung bei Kunden-Kommunikations-  
Dienstleistern ..... 371

*Bernd Juraschko*

Das Rechercherecht – Möglichkeiten und Grenzen..... 375

*Heiko Klarl*

Ausblick auf die modellgetriebene, mustergestützte Sicherheit in  
serviceorientierten Architekturen ..... 381

*Ralph Kölle*

Virtuelle Mitglieder in virtuellen Teams – Kompensation defizitärer  
Rollen durch Simulation Unterstützung virtueller Teams beim Lernen  
einer Programmiersprache..... 385

*Alexander Warta*

Quantitative Bewertung von Wiki-Kollaboration in der Industrie..... 391

*Hanna Knäusl*

Studentisches Programm der ISI 2007..... 397

## **„Form folgt Funktion?“**

### **Anforderungen, Wirklichkeit und Entwicklungsperspektive der Informationsversorgung in den Wissenschaften**

#### **13. Jahrestagung der IuK-Initiative Wissenschaft**

### **Information Retrieval in der Fachinformation**

*Manfred Hauer*

IntelligentCAPTURE und dandelon.com: Collaborative Catalog Enrichment... 403

*Peter Weiland, Andreas Gerards & Michael Gerards*

Einsatz automatischer Indexierungsverfahren in der Datenbank PSYINDEX. 413

*Siw Waffenschmidt*

Zwischenbericht aus der Virtuellen Fachbibliothek Sportwissenschaft..... 423

### **Informationen und Prozesse**

*Jutta Marx*

Information folgt Prozess. Prozessorientierte Informationssysteme zur Qualitätssicherung kleinerer und mittlerer Forschungseinrichtungen .....427

*Robert Roggenbuck*

MathematikerInnen machen sich nützlich: Projektinformationen und eine Expertendatenbank der angewandten Mathematik..... 433

*Susanne Pillay*

Change Management und Strukturveränderung bei IuK-Projekten an Universitäten 445

### **Elektronisches Publizieren**

*Verena Gondek*

Dokumentvorlagen zur Erstellung digitaler Hochschulschriften mit LaTeX.. 465

*Cordula Nötzelmann*

Die Initiative Digital Peer Publishing NRW ..... 469

*Björn Mittelsdorf & Ute Wahner*

Erfahrungen mit Community-Funktionen im PsychLinker ..... 479

### **Neue Entwicklungen in der Fachinformation**

*Sabine Heinz & Maximilian Stempfhuber*

Eine Informationsarchitektur für wissenschaftliche Fachportale in vascoda ... 485

*Stefan Baerisch*

Heterogenität in wissenschaftlichen Fachdatenportalen ..... 509

# **10. Internationales Symposium für Informationswissenschaft**

## **Veranstalter**

Hochschulverband Informationswissenschaft (HI), e. V.

## **Conference Chair**

Achim OSSWALD, Köln

## **Program Chair**

Christian WOLFF, Regensburg

## **Lokale Organisation**

Fachhochschule Köln, Institut für Informationswissenschaft

## **Programmkomitee**

Bernard BEKAVAC, HTW Chur

Nicholas BELKIN, Rutgers University

Christoph BLÄSI, Universität Erlangen-Nürnberg

Rafael CAPURRO, Fachhochschule Stuttgart

Elisabeth DAVENPORT, Napier University Edinburgh

Maximilian EIBL, TU Chemnitz

Wolfgang F. FINKE, Fachhochschule Jena

Simone FÜHLES-UBACH, Fachhochschule Köln

Hermann RÖSCH, Fachhochschule Köln

Norbert FUHR, Universität Duisburg-Essen

Ursula GEORGY, Fachhochschule Köln

Rainer HAMMWÖHNER, Universität Regensburg

Ilse HARMS, Universität Saarbrücken

Ralf-Dirk HENNINGS, Fachhochschule Potsdam

Josef HERGET, HTW Chur

Gerhard KNORZ, Fachhochschule Darmstadt

Traugott KOCH, Max Planck Digital Library, Berlin

Jürgen KRAUSE, IZ Sozialwissenschaften, Bonn

Otto KRICKL, Universität Graz

Rainer KUHLEN, Universität Konstanz

Franz LEHNER, Universität Passau

Heinz-Dirk LUCKHARDT, Universität Saarbrücken

Thomas MANDL, Universität Hildesheim

Achim OSSWALD, Fachhochschule Köln

Jiri PANYR, Siemens AG

Wolf RAUCH, Karl-Franzens-Universität Graz

Ulrich REIMER, Universität Konstanz

Harald REITERER, Universität Konstanz

Marc RITTBERGER, Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt

Christian SCHLÖGL, Karl-Franzens-Universität Graz

Ralph SCHMIDT, Hochschule für angewandte Wissenschaften Hamburg

Wolfgang SCHMITZ, Universitäts- und Stadtbibliothek Köln

René SCHNEIDER, HEG Genf

Wolfgang G. STOCK, Universität Düsseldorf

Peter SÜTHEÖ, Andrassy-Universität Budapest

Ulrich THIEL, Fraunhofer IPSI Darmstadt

Christian WOLFF, Universität Regensburg

Christa WOMSER-HACKER, Universität Hildesheim

Fahri YETIM, Fachhochschule Köln

Harald Zimmermann, Universität Saarbrücken

## **13. Jahrestagung der IuK-Initiative Wissenschaft**

### **Veranstalter**

IuK-Initiative Wissenschaft, e. V.

### **Program Chair**

Maximilian STEMPFHUBER, Bonn

### **Lokale Organisation**

GESIS – Informationszentrum Sozialwissenschaften, Bonn

### **Programmkomitee**

Frederik BORKENHAGEN, Deutsche Vereinigung für Sportwissenschaft, Hamburg

Norbert FUHR, Universität Duisburg-Essen

Detlef GÖRLITZ, Universität Hamburg

Maximilian STEMPFHUBER, IZ Sozialwissenschaften Bonn

Günter TÖRNER, Universität Duisburg-Essen

Erich WEICHSELGARTNER, Universität Trier

Luzian WEISEL, Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis, e.V. (DGI)





## Vorwort

Unter dem Motto „Open Innovation – neue Perspektiven im Kontext von Information und Wissen?“ sind in diesem Band Beiträge des 10. Symposiums für Informationswissenschaft und der 13. IuK-Tagung der wissenschaftlichen Fachgesellschaften versammelt. Die beiden Tagungen fanden 2007 – erstmals gemeinsam – an der FH Köln statt. Die Beiträge beleuchten daher Probleme der Informations- und Kommunikationstechnologien sowohl aus der Fachperspektive von Informationswissenschaft und Informatik als auch aus der Sicht wissenschaftlicher Fachgesellschaften. Zu den Themenschwerpunkten zählen

- *open access* und *open innovation*,
- Information Retrieval und Informationsvermittlung,
- Multimedia, Multimodalität und Multilingualität,
- Visualisierung in Informationssystemen,
- E-Learning und virtuelle Lernbedingungen,
- Wissensmanagement und insbesondere
- Kulturspezifisches Informationsverhalten und informationelle Vielfalt
- Digitalisierung von Kulturgütern
- Informationskompetenz für das Agieren in der digitalen Welt
- die heutige und zukünftige Rolle von *social software* im Kontext der Wissenschaftskommunikation.

Über 50 Beiträge aus Wissenschaft und Industrie mit Autoren aus sieben Ländern bieten eine breit gefächerte Sicht auf zentrale Probleme des Umgangs mit Information in der Wissenschaft.

Die Beiträge zur ISI 2007 wurden von mehr als 40 Fachgutachtern in weit über 100 Einzelgutachten geprüft, für alle Kurzbeiträge wurden wenigstens zwei, für die Langbeiträge wenigstens drei, z. T. bis zu fünf Gutachten erstellt. Die Ablehnungsquote bei den Kurz- und Langbeiträgen lag bei etwa einem Drittel der Einreichungen. Den Gutachtern sei für ihre Bereitschaft zur Mitwirkung sehr herzlich gedankt.

Im Programm der ISI sind studentische Beiträge ein wichtiger Bestandteil der ISI-Tagungen geworden. Mit der 10. ISI-Tagung in Köln wurden weitere Elemente integriert:

- Traditionsreich ist schon der Wettbewerb um den Gerhard-Lustig-Preis, mit dem die besten informationswissenschaftlichen Abschlussarbeiten der letzten beiden Jahre ermittelt und prämiert werden.
- Schon zum dritten Mal findet während des Internationalen Symposiums Informationswissenschaft ein studentisches Programm statt – in Köln erstmalig integriert in das offizielle Tagungsprogramm.
- Schließlich präsentieren – erstmalig in dieser Form – die Doktorandinnen und Doktoranden informationswissenschaftlicher Promotionen den Stand ihrer wissenschaftlichen Ausarbeitungen in Form von Postern und Kurzpräsentationen.

Alle drei Programmlinien drücken den hohen Stellenwert aus, den die Veranstalter den Ideen, der Kreativität und der Fachkompetenz der informationswissenschaftlichen Studierenden und Nachwuchswissenschaftlern zumessen.

Tagungen im Allgemeinen, aber erst recht in dieser Größenordnung mit mehreren Hundert Teilnehmern, können nur erfolgreich sein, wenn ein motiviertes Organisationsteam schon lange vor dem eigentlichen Konferenztermin beginnt, die notwendigen Vorbereitungen professionell mitzugestalten. Viele kleine und große Beiträge zum Gelingen einer solchen Veranstaltung finden dabei im Verborgenen statt. Um so mehr ist es uns ein Anliegen, den beteiligten Kolleginnen und Kollegen auf diesem Wege und vor der Fachöffentlichkeit für die gute Zusammenarbeit zu danken. Konkret gilt solcher Dank

- Sabine TRENKLER, Bettina ZACHARIAS und Patrick LAY vom Informationszentrum Sozialwissenschaften in Bonn
- Axel BADERSCHNEIDER und Manuel BURGHARDT, Universität Regensburg
- Simone FÜHLES-UBACH, Elke JUNG, Ursula GEORGY, Haike MEINHARDT, Isabella NAGY, Rusalka OFFER, Irena DE REUS, Ragna SEIDLER-DE ALWIS, Peter SLEEGERS, Inka TAPPENBECK und Fahri YETIM vom ISI-Organisationsteam des Instituts für Informationswissenschaft der Fachhochschule Köln
- der Leitung der Fachhochschule Köln sowie den beteiligten Mitarbeiterinnen und Mitarbeitern der Verwaltung der Fachhochschule Köln.

Unser Dank gilt darüber hinaus den Mitgliedern der verschiedenen studentischen Teams, die lange im Vorfeld und schließlich während der Tagung durch ihre Planungen und ihre tatkräftige Mithilfe u. a. auch ganz praktisch zur Durchführung der Veranstaltung beitrugen. Hier ist besonders Hanna KNÄUSL zu erwähnen, die das studentische Programm organisiert und für den Tagungsband beschrieben hat.

Schließlich ist Wolfgang SEMAR, Konstanz, herzlichst zu danken, der in bewährter Weise und – so unterstellen wir – in heiterer Gelassenheit die Produktion des Bandes begleitet und die Verbindung zu Verlag und Druckerei organisiert hat.

Köln, Bonn, Regensburg, im April 2007

Achim OSSWALD, Maximilian STEMPFHUBER, Christian WOLFF



# Florentine Coats of Arms on the Web:

## Experimenting with Retrieval Based on Text or Image Content\*

*Fabrizio Falchi<sup>1</sup>, Fausto Rabitti<sup>1</sup>,  
Werner Schweibenz<sup>2</sup> & Jan Simane<sup>2</sup>*

<sup>1</sup>ISTI-CNR, Pisa, Italy

<sup>2</sup>Kunsthistorisches Institut in Florenz Max Planck Institut, Florence, Italy

### Abstract

The paper describes two different ways of accessing a collection of Florentine coats of arms, which was digitised and made available on the Web. One way is a traditional approach: Textual description and indexing. As heraldry uses a specific and complex terminology, the textual approach works well for experts but is insufficient for laypersons. Therefore an additional retrieval mode was chosen. It is based on the visual specification by users of the content of images contained in the coat of arms. This access mode allows users who are not experts in heraldry to access the collection.

### 1 Introduction

The *Kunsthistorisches Institut in Florenz (KHI) Max-Planck-Institut* has possessed more than 2,800 coloured drawings of Florentine coat of arms since 1901. This unique collection is a very useful instrument for identifying coats of arms on palace facades, tomb monuments, altar pieces, and so on. The high importance and usefulness of this collection were the motives for the decision to digitise the material and make it accessible on the Web (the current version is a prototype and proof of concept, available online at <http://wappen.khi.fi.it/> and <http://www.nmis.isti.cnr.it/KHI/>). Another reason was to improve the access to the collection, which was indexed only rudimentarily on index cards. In addition to traditional textual indexing of the coats of arms, we wanted to try out retrieval based on image content because we considered the heraldic terminology to be a problem for laypersons. This claim can

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 1-13.

be illustrated by a fairly simple example of a blazon: “Gules a Griffin with Dragon Wings, Tail and Tongue rampant Or armed, beaked, langued and membered Azure between four Roses Argent.”

In this application context, we decided to explore the possibility of allowing the users to search the content of images representing the coat of arms using visual specification of this content. We believe that a very important move towards this kind of solution is to support *content-based retrieval* from the image database using *feature-based similarity access*. A feature (or *content-representative metadata*) is a set of characteristics of images, such as colour, texture and shapes. Similarity-based access means that the user specifies some characteristics of the wanted images, and the system retrieves the most relevant images with regard to the given characteristics, that is, the images *most similar* to the query. This approach assumes the ability to measure the distance (using metrics of some kind) between the query and the database images. This search paradigm:

Generalizes the information retrieval approach for textual archives, where search is based on key terms and the retrieved documents can be ranked with regard to the relevance of these key terms in the document text;

Generalizes the Web searching mechanism, as used by modern commercial systems, where searching is based on heuristics of links and domains and ranking is based on adapted text information retrieval techniques.

## 1.1 Description of the Material

The collection consists of two parts, on the one hand the coats of arms of Florentine families and on the other hand those of Florentine churches, hospitals and confraternities. The biggest part of the collection comprises the coats of arms of noble families, 2,512 colour drawings and 22 sketches. These drawings, which were done by an anonymous draughtsman in the 19<sup>th</sup> century, were acquired by the Institute library in 1901. A much smaller number of drawings refer to the coats of arms of Florentine churches, hospitals and confraternities. The 317 colour drawings were created by Otto Wenzel, the KHI’s librarian from 1902 to 1915, based on the Priorista di Luca Chiari.

In the library’s inventory of 1901, the drawings are described as “Sammlung von Wappen florentiner Familien u a.” [Collection of coats of arms of Florentine families and so on], while the other part is first mentioned in the inventory of 1909 as “Sammlung von Wappen florentiner Kirchen, Hospitälern und Bruderschaften

nach Priorista di Luca Chiari” [Collection of coats of arms of Florentine churches, hospitals and brotherhoods according to the Priorista di Luca Chiari].

Each coat of arms is drawn as a triangular shield in the centre of a sheet of light brown paper of approximately 16 x 23 cm; some of the sheets have watermarks. Most of the drawings are made in pencil and watercolour, a few of them are sketched only in pencil. Below each coat of arms, the name of the family is written in black ink. In some cases, commentaries in pencil are added. The whole collection is made accessible using a handwritten card index. This index combines the names of the families with the corresponding heraldic elements. For decades, both parts of the collections were stored in three boxes, one of them containing the index. During the 1980s, the sheets were stored in 65 black ring binders while the index cards remained in boxes.

The collection has been frequently consulted by scholars and researchers to identify unknown coats of arms on buildings and artworks or for looking up a certain family's coat of arms. In this way, it has become an important resource for art historians.

## **1.2 Reasons for the Digitisation**

For several reasons, the KHI decided to digitise this collection and make it accessible on the Web. Firstly, there were conservation issues because over time the drawings had become tattered to a certain extent due to extensive use. Therefore it seemed reasonable to digitise the material and provide access to digital facsimiles instead of the originals. Secondly, there were restrictions in physical access. The original drawings are kept in a special room with restricted office hours. The goal was to increase accessibility on-site and extend it to the public on the Web. The third reason was to improve access to the collection and the retrieval of individual coats of arms.

## **1.3 Project Description**

The project started in 2005. The first step was to provide a formal verbal description (blazon) of each individual coat of arms in German, according to the international standards of heraldry. As heraldry applies a complex specialist terminology, the KHI decided not to rely on heraldic manuals such as Spreiti [Spreiti 1928-1936] but to seek the advice of a specialist. Harald Drös, Head of the Research Institute for German Inscriptions at the Heidelberg Academy of Science, took on the responsibility of supervising this work. The information from the heraldic blazon for each coat of arms was entered in the KHI's art information database (HiDA). In the database, a variety of indices were created for relevant fields, such as family names,

colours, heraldic ordinaries and charges and so on, in order to allow combined searches for the content (see figure 1).

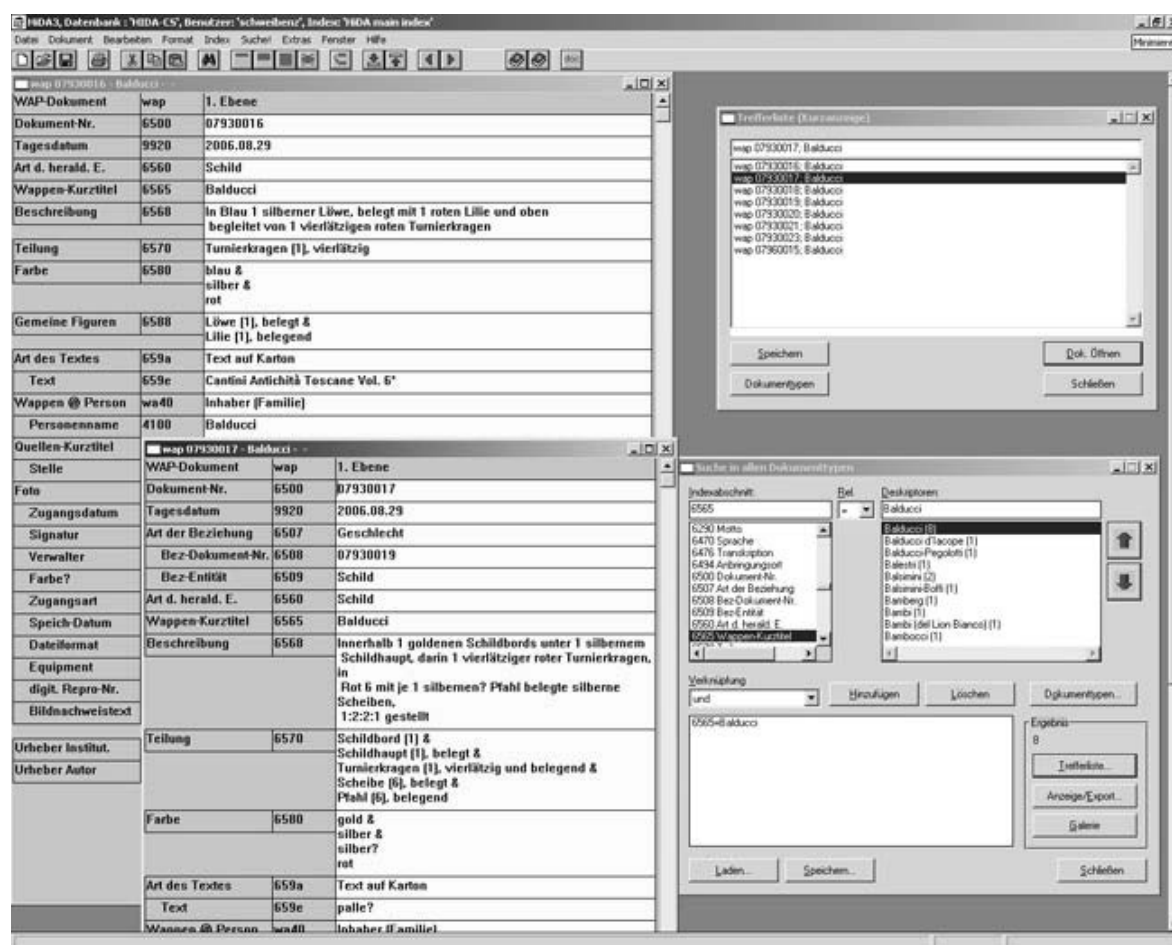


Figure 1: A heraldic blazon for a coat of arms in the HiDA database (Family name: Balducci; Heraldic charge: Lion)

In parallel, all the coats of arms were scanned in-house. The master scans were produced on a high-end scanner in RGB colour, with 16-bit colour depth and 600 pixels per inch. The height of each scan was approximately 5,500 pixels, the width approximately 4,000 pixels. This resulted in a file size of about 125 MB for the master. Each master was reduced to a working copy in lower quality with 8-bit colour depth and 300 pixels per inch for further use (for example, print). For the display on the Web, JPEGs with a width of 350 pixels are used.

The content of the art information database and the digitised images were integrated in an open-source Web database (ZOPE), which was provided by the KHI's technical partner, Fafalter GmbH in Düsseldorf. The same product has been successfully used for the KHI's digital photo library [Bieber, Schweibenz 2004, 2005]. In addition to the traditional text-based searches, the KHI decided to explore new means of access. The idea was to analyse the pictorial content of the individual coats



of arms and provide a query by image content. In order to implement this feature, the KHI looked for a partner with profound competence in this field. This was the *Istituto di Scienza e Tecnologie dell'Informazione/Consiglio Nazionale di Ricerca* (ISTI/CNR) in Pisa.

## 2 Text-Based Retrieval in the Coat of Arms Database

The text-based retrieval relies on the indices created in the art information database HiDA (Hierarchical Data Administrator) in German. By entering information in certain HiDA fields, a controlled vocabulary for different subjects is set up, for example, for tinctures (colours) such as “gold”, “silver” or “blue”, for heraldic ordinaries such as “bend”, “cross” or “chief” or for heraldic charges such as “lion” or “eagle”. Other fields in HiDA, such as the blazon information or the text information contained on the sheet of paper, are automatically indexed in full-text form. All indices are transferred to the ZOPE database and can be searched by entering a term in the search field on the Web interface (see figure 2) or in the advanced search (see figure 3).

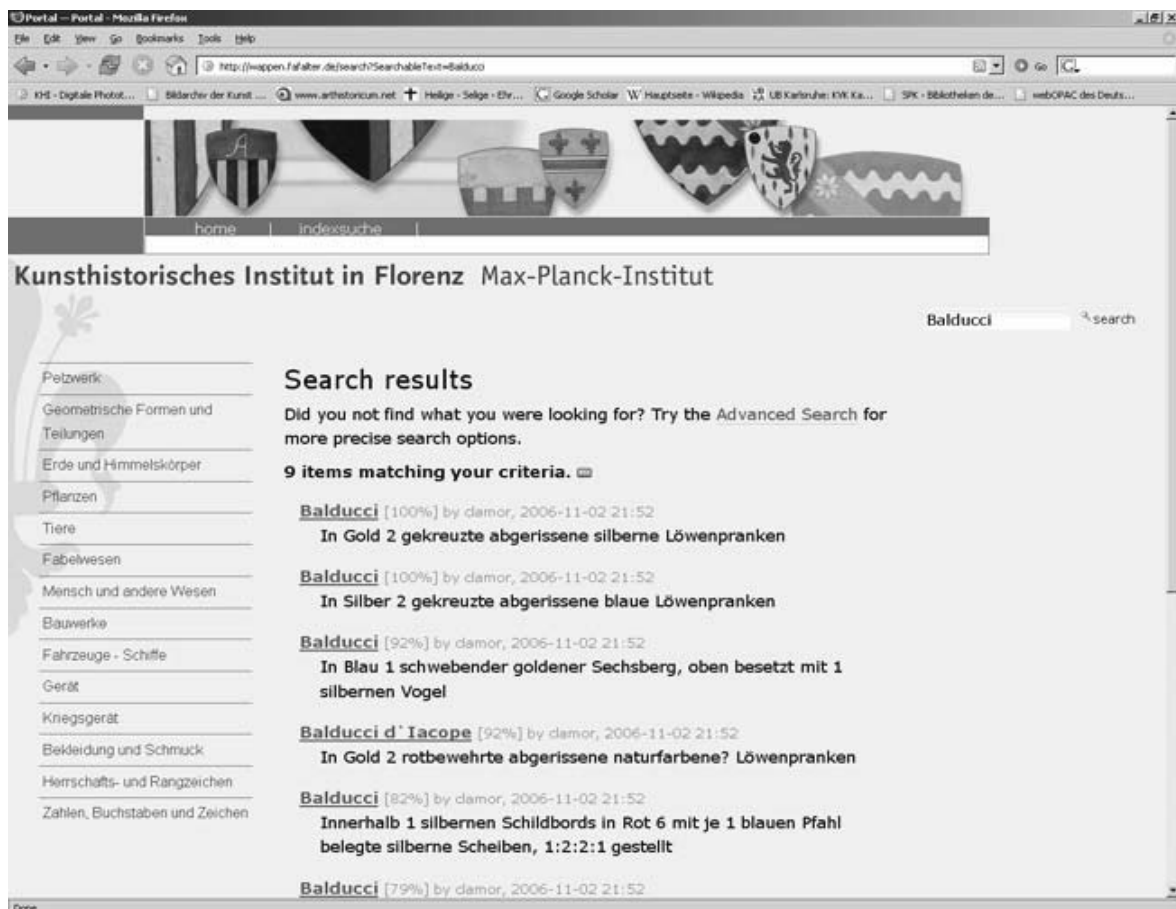


Figure 2: Live search for a heraldic charge (lion)

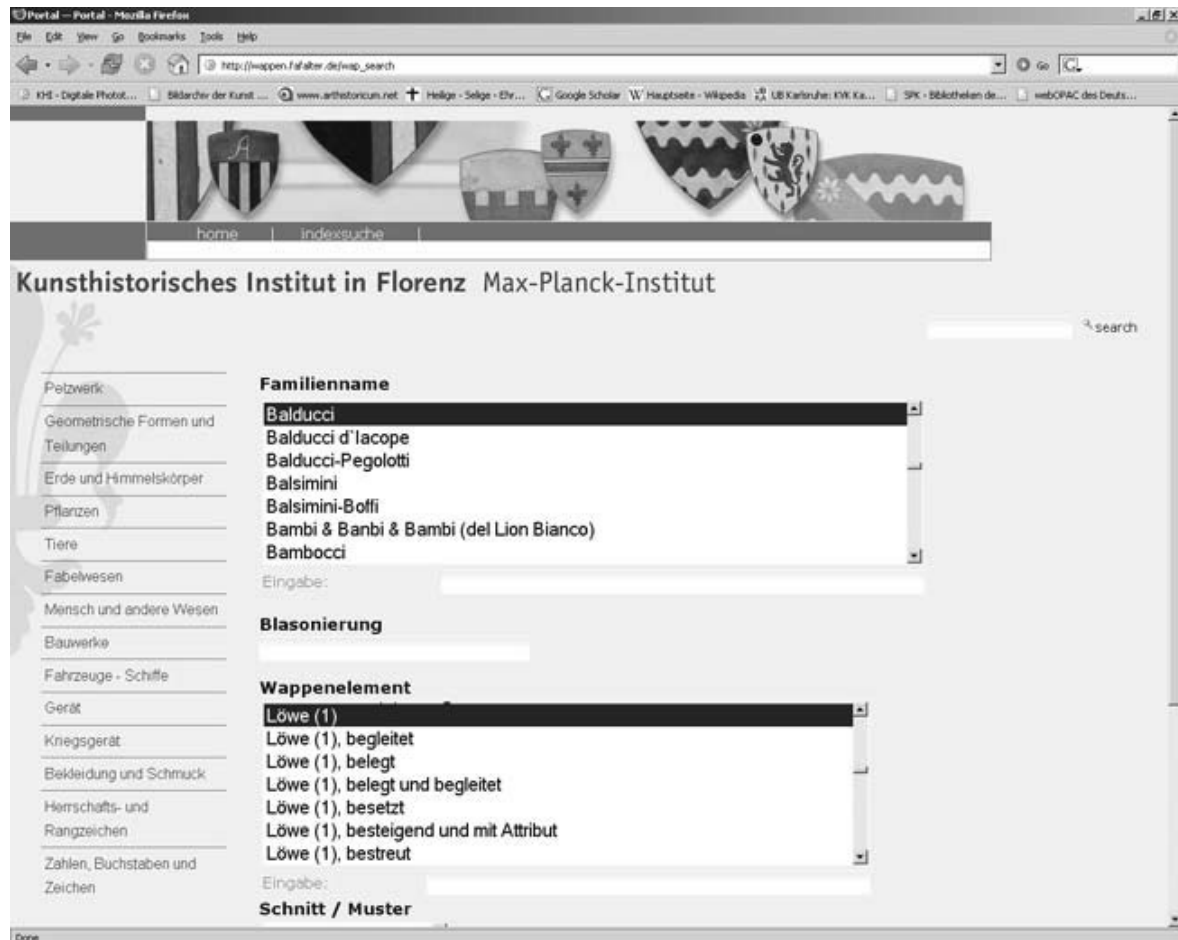


Figure 3: Combined search for a family name (Balducci) and a heraldic charge (lion)

As already mentioned above, the terminology of heraldry is very complex. Therefore we cannot expect the users to be familiar with all the specific terms. For this reason, apart from the simple and advanced search, a number of categories were added. These categories, based on a German heraldry book called “Wappenbilderordnung” [Wappenbilderordnung 1986], for example “animals,” “plants” or “mythical creatures”, allow users to approach the coats of arms starting from very general terminology and getting more and more specific.

Each category branches out, for example, animals into mammals and birds, and finally leads to a list of family names with an accompanying blazon. The categories are based on the indices provided by HiDA and are set up by the Content Management System PLONE using a service called “smart folders”. This allows users who are non-specialists to access the coats of arms in a convenient way without having to know or use heraldic terminology. Another means of access for non-specialists is the query by image content.

### 3 Content-Based Searching in the Coats of Arms Image Database

We had different options for implementing the retrieval application based on the visual specification by users of the content of images represented in the coat of arms.

The earliest and most common approach is the global query-by-example paradigm. It consists of retrieving images whose visual appearance is globally similar to a selected example image. Initially proposed by Swain and Ballard [Swain, Ballard 1999], it was adopted by a vast majority of content-based image retrieval systems (for example, IBM QBIC, VisualSEEk, Virage's VIR Image Engine and Excalibur's Image RetrievalWare). However, this paradigm has limited application in our case, since it would allow users to retrieve only coats of arms similar to the one specified.

The partial query-by-example paradigm was introduced later on. This approach allows the user to explicitly select a visual component which is relevant for the query and retrieves images which contain a similar visual component. This approach proved to be more selective, hence more precise than the global query-by-example paradigm [Ma, Manjunath 1999]. This is really what is needed for our application: Searching images of coats of arms based on a visual element which they contain.

In our approach, based on a similarity search of images, it is possible to apply the relevance feedback mechanism, inspired by text retrieval, to refine the image search [Huang, Rui, Mehrotra 1997]. Among the retrieved images, the user specifies which ones are relevant and reiterates the search. By refining the similarity measure, the searched image can be reached more efficiently, taking into account the subjective preferences of the user.

A fundamental decision regarding the approach taken is the adoption of the MPEG-7 standard to represent the visual metadata, that is, the features extracted from the coat of arms images. MPEG-7 [ISO/IEC 15938], also called the Multimedia Content Description Interface, aims to cover the need for searching and retrieving multimedia information by describing the content of audiovisual objects in a standardised way. The MPEG-7 approach is to specify a set of standardised *descriptors*, that is, entities that contain the syntax and semantics of audiovisual content. For example, a descriptor could refer to the colour of an object by specifying the name of the attribute colour (such as, Color) and the type of the value (such as, a string or three integer values for RGB colours).

### 3.1 Image Analysis and Segmentation

Before using the images of the database in the retrieval process, it is necessary to analyse their content and identify the relevant components. This process is called segmentation.

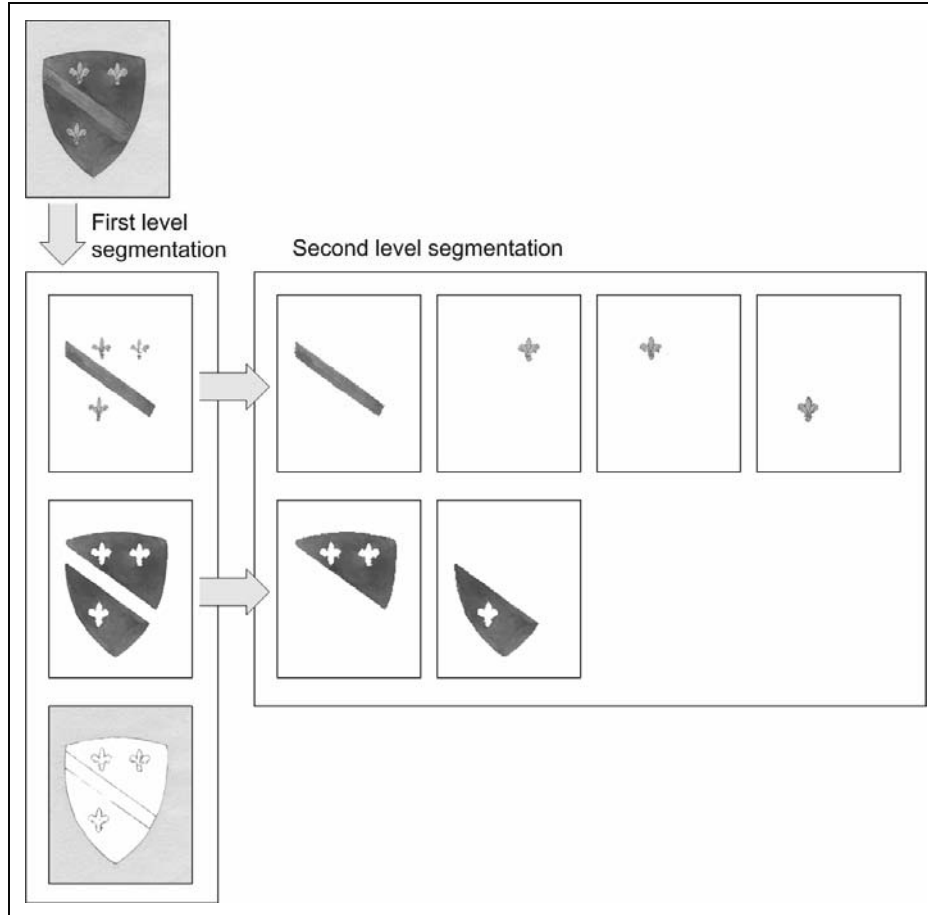


Figure 4: Two levels of segmentation

Image segmentation refers to partitioning the image into homogeneous regions. We decided to use two levels of segmentation using two different segmentation algorithms. Basically, the first one divides the image into regions based on their colour gradient. The results of this first-level segmentation are not necessarily connected regions of similar colour. These regions are then segmented again using another algorithm based on K-Means, which is an algorithm to cluster objects based on attributes trying to minimize total intra-cluster variance. The results of this second-level segmentation are connected regions. All regions obtained by both segmentation algorithms are used.

For the first-level segmentation, we used an implementation of the algorithm described in [Comaniciu, Meer 1997], developed by Dorin Comaniciu. The tech-

nique is based on the mean shift algorithm, a simple nonparametric procedure for estimating density gradients.

For the second-level segmentation, we used a novel variant of the well known K-Means-with-connectivity-constraint algorithm (KMCC), a member of popular K-Means family, which is described in [Mezaris, Kompatsiaris Strintzis 2002, 2004]. The KMCC algorithm classifies the pixels into regions taking into account not only the intensity information associated with each pixel but also the position of the pixel, thus producing connected regions rather than sets of chromatically similar pixels.

### 3.2 MPEG-7 Descriptors Extraction and Searching

Feature extraction was performed using an application we built based on the MPEG-7 [ISO/IEC 15938] experimentation model [ISO/IEC 15938-6:2003] of MPEG-7. The software can extract all MPEG-7 image visual descriptors [ISO/IEC 15938-3:2003]. The same application has been used on MILOS [MILOS, Amato, Gennaro, Rabitti, Savino 2004] and particularly on PhotoBook [Amato, Bolettieri, Debole, Falchi, Rabitti, Savino 2006, Amato, Bolettieri, Debole, Falchi, Rabitti Savino 2006], which is a Multimedia Digital Library Application (<http://milos.isti.cnr.it>) that we built on top of MILOS for online photo sharing.

For both complete coat of arms images and automatically selected regions, we extracted the following MPEG-7 descriptors: *ScalableColor*, *ColorStructure*, *ColorLayout*, *DominantColor*, *EdgeHistogram* and *HomogeneousTexture*.

Because of the fact that they are concerned with shape, 2 MPEG-7 descriptors were extracted only for regions: *RegionShape* and *ContourShape*.

Two more descriptors for colour were used for complete images: *ColorStructure* and *ColorLayout*. These descriptors were not used for regions because they are meaningless for non-rectangular images.

We now provide a brief description of the MPEG-7 descriptor used. *ScalableColor*, *ColorStructure*, *ColorLayout* and *DominantColor* are all to do with colours. *ScalableColor* is a colour histogram in the HSV Colour Space. *ColorStructure* captures both colour content and information about the spatial arrangement of the colours. *ColorLayout* represents the spatial layout of colour images. *EdgeHistogram* describes the spatial distribution of five types of edges. *HomogeneousTexture* characterizes the properties of texture in an image. *RegionShape* expresses pixel distribution within a region. *ContourShape* is based on the Curvature Scale-Space representation of the contour.

The result of the extraction process is an XML document. The online application developed for this project performs a linear scan of the descriptors present in the document. Given the limited number of images in the database, just a few thousand, the use of an index structure for the similarity search is not necessary and a linear search of the MPEG-7 descriptors is feasible. However, in MILOS we use M-Tree [Salembier, Sikora 2002] as an index structure for the similarity search: This allows the system to be scalable, searching tens of million of MPEG-7 image descriptors in a few seconds.

To compare the descriptors resulting from two images, we used the functions proposed in [Swain, Ballard 1991]. If the user wants to search using a combination of descriptors, distances are combined, single-feature results are accumulated using predefined (heuristic) weights.

### 3.3 Online Image Content Search Application

Once an image is selected (for example, from the textual description database), the application asks the user to select the region (or even the complete image) that he or she wants to use as a query to search for similar ones. The first 10 results are listed in order of decreasing similarity using *EdgeHistogram* and *RegionShape*, which we found to be the most useful descriptors for this kind of images (see fig. 5).

From the results page, it is possible to select the descriptors to be used for the search and then ask for the relative results. When the user clicks a result region, the search is performed again using the selection as the query region.

Below each results region, there is a small image of the entire coat – of arms which the region comes from. Using the *Info* link located above each region present on the results page, the user can access the textual metadata in the database. On the upper left side, the results page also reports the query coat-of-arms and the region used as the query.

## 4 Conclusion and Future Research

The project combines two different ways of accessing a collection of Florentine coats of arms, one being traditional and text-based, the other being experimental image searching-by-content. With this combination, we hope to satisfy the needs of both experts and laypersons in the field of heraldry as experts can use the heraldic terminology for a text-based search while laypersons can use the content-based search.

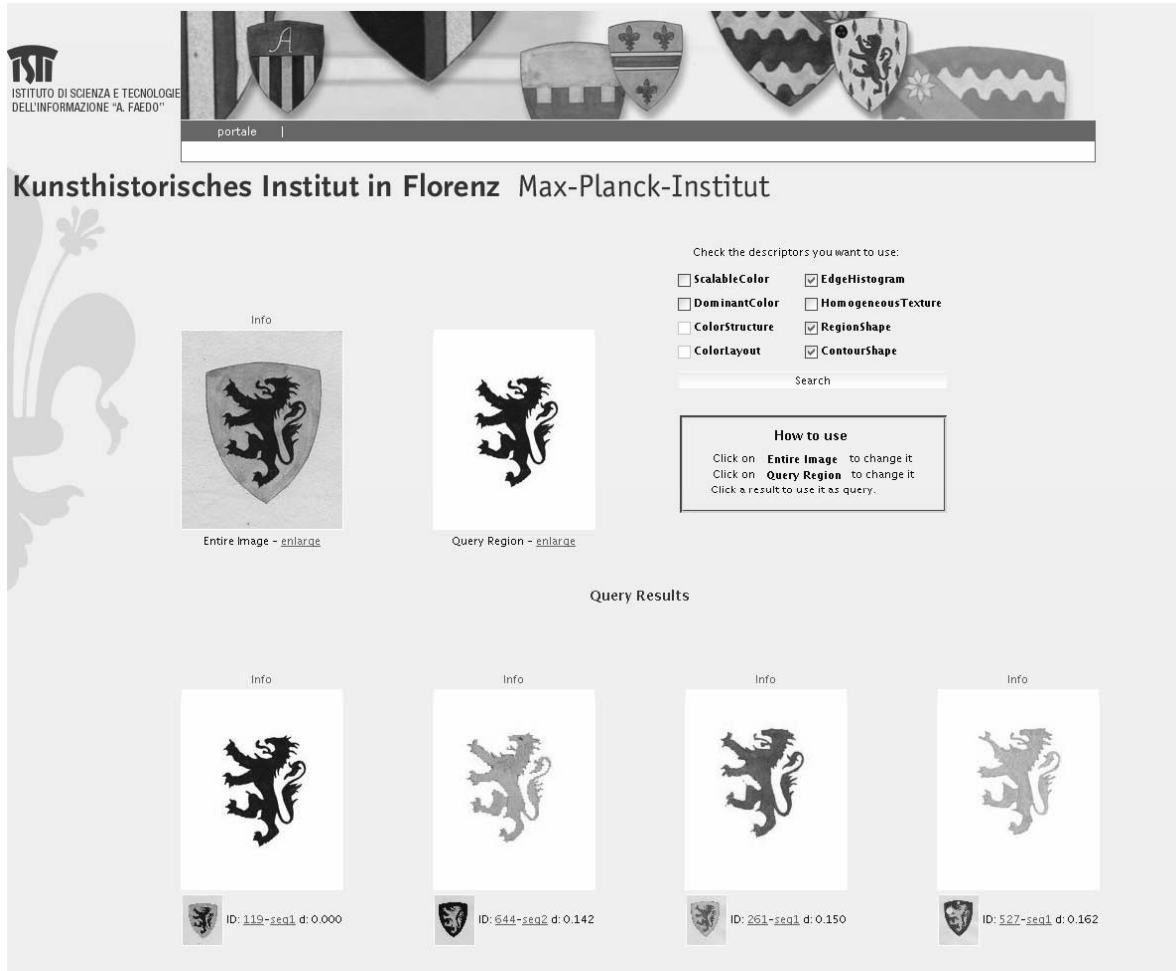


Figure 5: Search interface

The text-based approach is based on controlled vocabulary from heraldic vocabulary (colours, ordinaries, charges, and so on) and vocabulary from free-text fields such as blazon information and text information. The index search allows users to combine entries in selected fields, such as family name, blazon, ordinaries and charges. To support the users, the correct terminology for the fields is shown in a scroll field. Nevertheless, we think it difficult for non-experts to practice efficient retrieval based on text alone, due to the complex nature of the heraldic terminology which is only available in German.

An alternative approach, which does not need terminological knowledge or knowledge of the German language on the part of the users, is an image-based search. We have described the implementation of an image searching-by-content application in which a user can retrieve images of coats of arms by indicating some of their visual components. We will have to investigate how this approach is accepted by users, on the basis of future experience and compared to the text-based approach. In any case, the main advantage of this approach is that the procedure (segmentation, MPEG-7

feature extraction, content indexing) is completely automatic, once the images are digitised, and so can be applied to huge image databases with a minimum of effort.

## **5 Bibliography**

- [Amato, Gennaro, Rabitti, Savino 2004] G. Amato, C. Gennaro, F. Rabitti, and P. Savino. Milos: A multimedia content management system for digital library applications. In European Conference on Digital Libraries, ECDL 2004, Bath, UK, September 12-17 2004, 2004.
- [Amato, Bolettieri, Debole, Falchi, Rabitti, Savino 2006] G. Amato, P. Bolettieri, F. Debole, F. Falchi, F. Rabitti and P. Savino: Using MILOS to Build a Multimedia Digital Library Application: The PhotoBook Experience. ECDL 2006: 379-390
- [Amato, Bolettieri, Debole, Falchi, Rabitti Savino 2006] G. Amato, P. Bolettieri, F. Debole, F. Falchi, F. Rabitti and P. Savino: Using MILOS to build an on-line photo album: the PhotoBook. SEBD 2006: 233-240
- [Bieber, Schweibenz 2005] C. Bieber and W. Schweibenz. Images of Italian Art on the Web. The Database of the Photo Library of the German Art Institute in Florence. In: New Library World 106 (9/10) 2005. 430-440.
- [Bieber, Schweibenz 2004] C. Bieber and W. Schweibenz. Digitalisierung und Langzeitarchivierung von Fotonegativen. In: Information zwischen Kultur und Marktwirtschaft. Proceedings des 9. Internationalen Symposiums für Informationswissenschaft (ISI 2004) Chur 6.-8. Oktober 2004. Hrsg. von Bernard Bekavac, Josef Herget, Marc Rittberger. (Schriften zur Informationswissenschaft 42). Konstanz: UKV. 125-140.
- [Comaniciu, Meer 1997] D. Comaniciu and P. Meer. Robust Analysis of Feature Spaces: Color Image Segmentation. IEEE Conf. on Comp. Vis. and Pattern Recognition, pages 750--755, Puerto Rico 1997.
- [Huang, Rui, Mehrotra 1997] T. Huang Y. Rui and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. IEEE International Conference on Image Processing (ICIP), 1997.
- [ISO/IEC 15938] ISO/IEC. Information technology – Multimedia content description interfaces. 15938.
- [ISO/IEC 15938-6:2003] ISO/IEC. Information technology – Multimedia content description interfaces. Part 6: Reference Software. 15938-6:2003.
- [ISO/IEC 15938-3:2003] ISO/IEC. Information technology -Multimedia content description interfaces. Part 3: Visual. 15938-3:2002.
- [Ma, Manjunath 1999] W. Y. Ma and B. S. Manjunath. Netra: A toolbox for navigating large image databases. Multimedia Systems, 7(3):184{198, 1999.
- [Mezaris, Kompatsiaris Strintzis 2004] V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "Still Image Segmentation Tools for Object-based Multimedia Applications", International Journal of Pattern Recognition and Artificial Intelligence, vol. 18, no. 4, pp. 701-725, June 2004.



- [Mezaris, Kompatsiaris Strintzis 2002] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "A Framework for the Efficient Segmentation of Large-format Color Images", Proc. IEEE International Conference on Image Processing (ICIP 2002), vol. 1, pp. 761-764, Rochester, NY, USA, September 2002.
- [MILOS] MILOS: Multimedia dIgital Library for On-line Search. <http://milos.isti.cnr.it/>.
- [Salembier, Sikora 2002] P. Salembier and T. Sikora. Introduction to MPEG-7: Multimedia Content Description Interface. John Wiley & Sons, Inc., New York, NY, USA, 2002.
- [Spreti 1928-1936], Enciclopedia storico-nobiliare italiana: famiglie nobili e titolate viventi, riconosciute dal R°. Governo d'Italia / promossa e diretta dal Marchese Vittorio Spreti, 1928-1936.
- [Swain, Ballard 1991] M. Swain and D. Ballard. Color indexing. International Journal of Computer Vision (IJCV), 7(1):11-32, 1991.
- [Wappenbilderordnung 1986] Wappenbilderordnung. Symbolorum Armorialium Ordo, hrsg. vom Herold, Verein für Heraldik, Genealogie und verwandte Wissenschaften zu Berlin. Bearb. von Jürgen Arndt und Werner Seeger. Neustadt an der Aisch: Bauer & Raspe, 1986.



# Geographische Bedingungen im Information Retrieval:

## Neue Ansätze in Systementwicklung und Evaluierung\*

*Kerstin Bischoff, Ralph Kölle,  
Thomas Mandl & Christa Womser-Hacker*

Informationswissenschaft  
Universität Hildesheim  
Marienburger Platz 22, D-31141 Hildesheim  
*mandl@uni-hildesheim.de*

### Zusammenfassung

Anfragen an Information Retrieval Systeme beinhalten häufig geographischen Einschränkungen. Solche geographischen Angaben wie Städte- oder Ländernamen ermöglichen Suchmaschinen die Optimierung ihrer Ergebnisse jenseits von klassischen Indexierungstechniken. Beispielsweise lassen sich geographische Entitäten in Ontologien einordnen und liefern somit gute Hinweise für Erweiterungsterme. Dieser Beitrag stellt den Stand der Forschung dar, analysiert aktuelle Evaluierungsergebnisse und leitet daraus Optimierungsansätze für die Systementwicklung ab. Experimente und deren Auswertung zeigen, dass flexible Anwendungen von Relevanz Feedback zu verbesserten Ergebnissen führen können.

### Abstract

Queries with a geographic scope are often posted to information retrieval systems. Geographic search terms conditions like city or country names allow search engines to optimize their results beyond classic indexing techniques. Geographic entities can be identified within ontologies and be properly extended, for example. This paper briefly reviews the state of the art, analyses current evaluation results and presents new optimization techniques. Experiments show that the flexible application of relevance feedback can improve the retrieval results.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 15-26.

## I Einleitung

Internet-Suchmaschinen gehören für viele Menschen bereits zum Alltag und bieten den wichtigsten Zugang zum Wissen im Internet. So werden nach Schätzungen täglich mehr als 500 Millionen Anfragen an Suchmaschinen gestellt<sup>1</sup>, um unter Milliarden von Seiten im Internet die relevantesten auszusuchen. Die grundlegenden Verfahren für Suchmaschinen liefert das Information Retrieval, das heute eine der Schlüsseltechnologien in der Wissensgesellschaft darstellt. Information Retrieval behandelt die Repräsentation des Inhalts der Dokumente und den Abgleich zwischen Anfragen des Benutzers und den gespeicherten Repräsentationen.

Zahlreiche Anfragen an Suchmaschinen enthalten geographische Begriffe. Informationen müssen nicht nur inhaltlich relevant sein, sondern sollen sich auch konkret auf den angegebenen Ort beziehen. Diese geographischen Anfragen lassen sich vielfältig klassifizieren. Bei Eingaben wie „*Apotheke mit Nachtdienst in Nandlstadt*“ wünschen Benutzer einzelne Seiten mit konkreten Informationen, während „*Investitionen in Hastenbeck*“ auf eine vorab nicht festlegbare Menge von Dokumenten zu einem bestimmten Thema abzielt. Erstere werden im Web-Retrieval häufig als *navigational queries (known item search)* und die zweite Gruppe als *informational queries* (klassisches ad-hoc Retrieval) bezeichnet [Broder 2002].

Geographische Eigennamen können wie andere Eigennamen das Retrieval verbessern. Einige Suchmaschinen liefern teilweise bereits Ergebnisdokumente, welche zu Landkarten führen. Darüber hinaus bieten die geographischen Begriffe weitere Möglichkeiten zu formalen Schlüssen. Liegen Ontologien vor, so kann beispielsweise automatisch mit geeigneten Unterbegriffen erweitert werden (z. B. *Oberpfalz* -> *Nabburg, Wackersdorf, ...*).

Andererseits stellen Suchanfragen mit geographischen Eigennamen auch besondere Herausforderungen an die Anfragenverarbeitung. Die Erkennung und Klassifikation von Eigennamen gelingt heute noch nicht zu 100%. Der Gebrauch für politische als auch geographische Einheiten (*Europa*), metaphorischer Einsatz (*Tal der Ahnungslosen*), vage geographische Bezeichnungen (*Mittelmeerraum, Ostbayern, Tropen*) oder Mehrwortgruppen (*Conselheiro Lafaiete, Neunburg vorm Wald*) seien als Beispiele für die Probleme bei der Erkennung und Verarbeitung genannt. Gerade in mehrsprachigen Kontexten verschärfen sich die Schwierigkeiten für die Systeme. Manche geographischen Eigennamen müssen übersetzt oder zerlegt werden (z. B.

---

<sup>1</sup> <http://searchenginewatch.com>

Deutsch: *Donau*, Spanisch: *Danubio*, Ungarisch: *Duna*, etc.), andere müssen in ihrer Originalform bestehen bleiben (z. B. *Waldheim*).

## 2 Geographisches Information Retrieval

Mit den heute verfügbaren Technologien kann die Bedeutung von Texten (in Anwendungen für Massendaten) nicht zu 100% erkannt und repräsentiert werden. Die Systeme begnügen sich daher mit unsicheren Repräsentationen (*bag of words*), welche die Bedeutung nur vage widerspiegeln [Mandl & Womser-Hacker 2006]. Geographische Informationssysteme erlauben dagegen exakte Suchen, da geographische Beziehungen exakt im Sinne des Datenbank-Retrieval repräsentiert werden. Im geographischen Information Retrieval treffen diese beiden scheinbar widersprüchlichen Paradigmen zusammen. Für geographisches Information Retrieval wurden zahlreiche plausible Optimierungsansätze vorgeschlagen, wozu besonders das Hinzuziehen von Wissensressourcen und Schlussfolgerungen zählen. Das Zusammenspiel mehrerer Komponenten in einem System und ihre Effektivität in Bezug auf ein bestimmtes Korpus lassen sich jedoch nie vorhersagen. Somit kommt der Evaluierung der Effektivität auch beim geographischen Retrieval entscheidende Bedeutung bei.

Umfangreiche Evaluierungen und Systemvergleiche erfolgen heute im Rahmen großer Evaluierungsinitiativen [Mandl 2006], welche eine Infrastruktur für die Evaluierung von Systemen mit Massendaten zur Verfügung stellen. Diese Infrastruktur umfasst Dokumenten-Korpora, Aufgaben für die Systeme (*Topics*), intellektuelle Relevanz-Bewertungen sowie die statistische Auswertung. Als Trainingsdaten stehen den teilnehmenden Systemen dabei die Korpora der Vorjahre zur Verfügung, die Relevanzurteile enthalten. Den Besonderheiten des mehrsprachigen Retrieval widmet sich seit dem Jahr 2000 das Cross-Language Evaluation Forum (CLEF<sup>2</sup>, cf. [Peters et al. 2006]). Die CLEF-Organisatoren bieten eine mehrsprachige Infrastruktur für alle beteiligten Sprachen. *Cross Language* Information Retrieval (CLIR) modelliert die Situation von Benutzern, die eine Fremdsprache zwar passiv beherrschen und die Relevanz von Dokumenten in dieser Sprache zumindest abschätzen können, deren aktive Sprachkenntnisse aber nicht für die Konstruktion einer guten Anfrage ausreichen. Mehrsprachiges und insbesondere sprachübergreifendes Information Retrieval erfordern weitere linguistische Komponenten.

---

<sup>2</sup> <http://www.clef-campaign.org>

Im Rahmen von CLEF wurde das geographische Retrieval als wichtige Herausforderung erkannt und in die Evaluierungsinitiative aufgenommen. Ein *Pilot Task* 2005 ging in einen regulären Task GeoCLEF 2006 über<sup>3</sup>. Die Gestaltung der Evaluierungsinfrastruktur fällt bei geographischem Retrieval schwer. Was sind für die bereitstehende Kollektion natürliche Anfragen mit geographischem Bezug? Wie fein sollen die geographischen Einheiten modelliert werden? Sollen explizite geographische Bezüge angegeben werden oder sollen Systeme diese selbst erkennen? In GeoCLEF 2005 gaben explizite geoTags im Markup der Anfrage klare Hinweise, während in GeoCLEF 2006 diese Zusatzhilfe wegfiel, nicht zuletzt weil sie wenig genutzt wurde. Die Verarbeitungsstrategie eines Systems hängt stark davon ab, ob einfache Suchwortübereinstimmungen geprüft oder aufwendige geographische Verarbeitungsschritte eingesetzt werden.

<top><num>GCo36</num>

<DE-title>Automobilindustrie rund um das Japanische Meer</DE-title>

<DE-desc>Küstenstädte am Japanischen Meer mit Automobilindustrie oder -werken</DE-desc>

<DE-narr>Relevante Dokumente berichten von Automobilindustrie oder -werken in Städten an der Küste des Japanischen Meeres (auch Ostmeer (von Korea) genannt), einschließlich wirtschaftlicher oder sozialer Ereignisse wie geplante Joint Ventures oder Streiks. Neben Japan grenzen auch die Länder Nordkorea, Südkorea und Russland an das Japanische Meer.</DE-narr></top>

*Abbildung 1: Beispiel für ein Topic aus GeoCLEF 2006*

GeoCLEF übernahm das in CLEF entwickelte, mehrsprachige, aus Zeitungsartikeln bestehende Korpus. Die Topics werden dabei parallel für unterschiedliche Sprachen entwickelt, indem möglichst realistische Benutzeranfragen modelliert, recherchiert und dann von Muttersprachlern in die jeweilige Sprache übersetzt werden. Im Jahre 2005 bot GeoCLEF 25 Topics in den Ausgangssprachen Deutsch, Englisch, Spanisch und Portugiesisch an. Die Testkollektionen standen in deutsch und englisch zur Verfügung, so dass monolinguale und bilinguale Experimente möglich waren. 2006 waren auch monolinguale Versuche (*runs*) gegen Kollektionen in Spanisch und Portugiesisch möglich, hinzu kamen Topics in Japanisch.

---

<sup>3</sup> <http://ir.shef.ac.uk/geoclef/>

### 3 Ergebnisse von GeoCLEF 2006

An GeoCLEF nahmen 17 Gruppen teil und lieferten 149 vollständige Experimente für alle 25 Topics ab [Gey et al. 2007]. Nicht zuletzt die Beteiligung von Microsoft, das beweist die hohe Relevanz der Thematik.

Für diese Untersuchung wurden die in Tabelle 1 angegebenen Runs ausgewertet. Als Vergleichswert bietet Tabelle 2 einige Ergebnisse von ad-hoc Runs.

<i>Run Typ</i>	<i>Anzahl Teilnehmer</i>	<i>Anzahl Runs</i>	<i>MAP top System</i>	<i>StAbw. Topics (absolut)</i>	<i>StAbw. Systeme (abs.)</i>
Monolingual DE	4	16	22,29%	16,07%	5,56%
Monolingual EN	16	73	30,34%	15,47%	7,84%
Bilingual X->DE	3	11	16,82%	18,00%	2,98%
Bilingual X->EN	3	12	27,07%	23,98%	4,18%

Tabelle 1: Ausgewertete Experimente GeoCLEF 2006 mit Mean Average Precision (MAP) und Standard-Abweichung (StAbw) der MAP zwischen den Topics und Systemen

<i>Run Typ</i>	<i>CLEF Track</i>	<i>Anzahl Runs</i>	<i>MAP top System</i>	<i>StdAbw. Topics (absolut)</i>	<i>StdAbw. Systeme (absolut)</i>
Bilingual X->EN	CLEF 2002 Ad-hoc	51	49,35%	24,98%	9,11%
Monolingual DE	CLEF 2002 Ad-hoc	21	52,34%	24,44%	11,45%

Tabelle 2: Vergleichsexperimente CLEF 2002 Ad-hoc Retrieval

In beiden Tabellen zeigt sich deutlich, dass die Topics für die höchste Varianz in den Ergebnissen verantwortlich sind. Dem gegenüber ist der Einfluss der Systeme eher gering. Im Vergleich zu den ad-hoc Experimenten sind die durchschnittlichen Precision-Werte (*mean average precision* = MAP) bei geographischem Retrieval weitaus niedriger. Dies kann bedeuten, dass für die Systeme geographisches Retrieval noch schwieriger ist oder dass die Anfragen aus GeoCLEF noch ungleich schwieriger sind als die ad-hoc Topics. Zwar gibt es keine generell akzeptierte Definition von einem schwierigen Topic [Grivolla & Mori 2005], innerhalb dieses Beitrags wird die Schwierigkeit als umgekehrt proportional zu der Systemleistung (MAP) verstanden.

Zwar sind die absoluten Werte in der Standardabweichung zwischen den besten Systemleistungen für die Topics in beiden Tabellen vergleichbar, aber die relative Varianz ist bei den geographischen Anfragen deutlich höher. Innerhalb der geographischen Anfragen bei CLEF liegen also erhebliche Unterschiede vor. Ein genauere Blick auf die Performanz der einzelnen Systeme soll erweisen, wie dieses Phä-

nomen zustande kommt. Die Retrieval-Qualität der besten zehn Systeme für zehn Anfragen verdeutlicht Abbildung 1.

Visuell zeigen sich drei Gruppen von Topics. Für einige Topics erzielen alle Systeme konstant Werte von über 65% MAP. Für eine größere Gruppe von Topics liegen die Werte aller Systeme unter 50% MAP. Bei Betrachtung aller 25 Topics tritt bei etwa fünf bis acht Topics eine starke Abweichung zwischen den Systemen auf. Die verschiedenen Systemleistungen (MAP) für die Topics variieren entweder innerhalb der zweiten Gruppe stark oder die Werte schwanken noch stärker. Für eines der Topics mit sehr unterschiedlichen Ergebnissen liefert selbst das beste System einen Wert von Null, während das zweitbeste System als einziges eine befriedigende Lösung erbringt, die dann aber gleich bei 100% liegt. Der sechs-beste Run in der Abbildung weicht deutlich von den anderen ab und liefert fast für jedes Topic andere Werte als vergleichbare Runs.

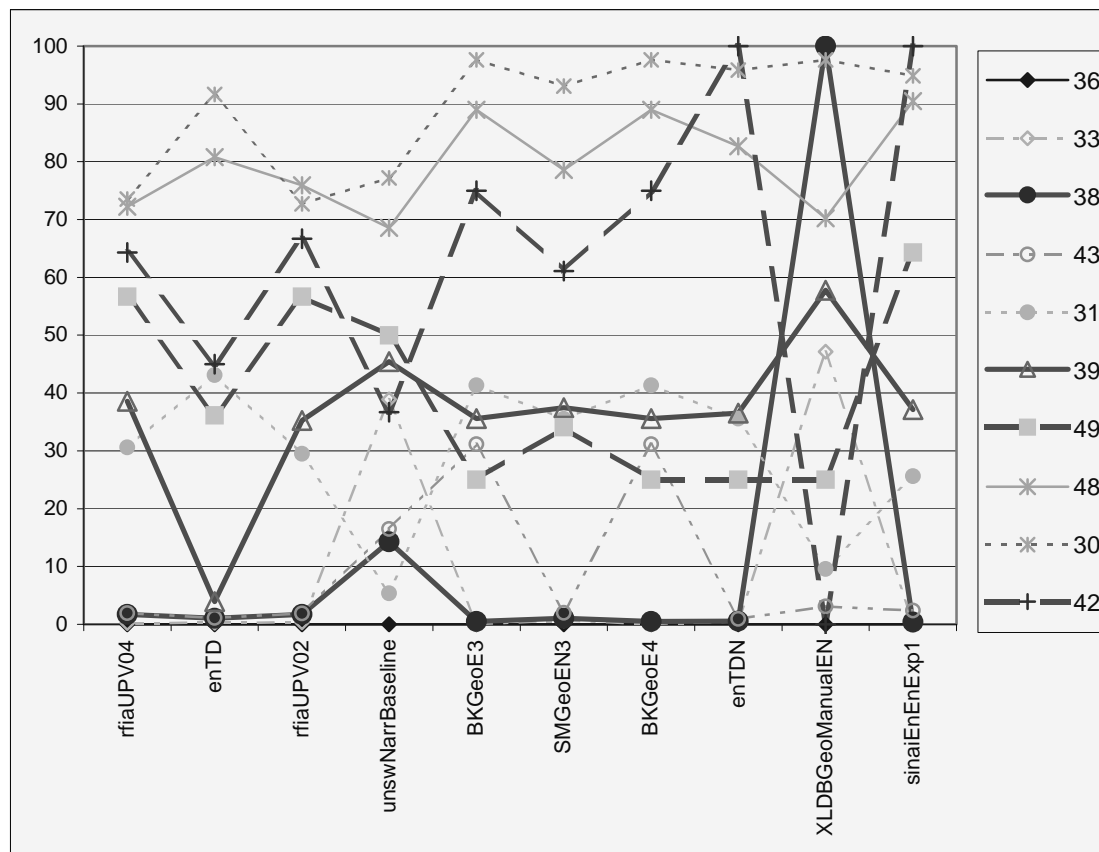


Abbildung 1: Unterschiedliche Leistung der besten zehn Systeme für zehn Topics

Um zu prüfen, welche Faktoren zu dieser hohen Varianz führen, erfolgte eine Analyse der Topics. Dazu wurde intellektuell analysiert, wie viele geographische Eigennamen jedes Topic enthielt und ob es sich um Länder- oder Kontinent-Be-



zeichnungen handelt. Die Anzahl der übrigen inhaltstragenden Begriffe ging ebenfalls in die Analyse mit ein. Für die Anzahl der geographischen Begriffe ergab sich keine klare Tendenz hinsichtlich der Auswirkung auf die Systemleistung. Alle Korrelationen zwischen der Performanz der Systeme für ein Topic und der Anzahl der untersuchten Begriffe im Topic (Länder-Namen, Kontinent-Namen und andere, nicht geographische, inhaltstragende Begriffe) zeigt Tabelle 3. Die Performanz wurde einmal zum besten System für jedes Topic (top) und einmal zum Durchschnitt aller Systeme für ein Topic (ave) berechnet.

	<i>Länder-Namen</i>		<i>Kontinent-Namen</i>		<i>Inhaltstragende Begriffe</i>	
	Top	Ave	Top	Ave	Top	Ave
Monolingual DE	0,23	0,08	-0,22	-0,18	0,17	0,18
Monolingual EN	0,14	0,27	0,05	-0,21	0,29	0,05
Bilingual X-> DE	0,28	0,16	-0,06	-0,09	0,15	0,10
Bilingual X-> EN	0,11	0,06	0,12	-0,11	0,27	0,10

Tabelle 3: Korrelationen zwischen MAP und in den Topics vorkommenden Begriffe für analysierte Experimente aus GeoCLEF 2006 für Deutsch (DE) und Englisch (EN)

Sämtliche Korrelationen sind sehr schwach. Lediglich für inhaltstragende Begriffe sowie für Ländernamen ergeben sich konsistente Tendenzen für alle vier Tracks. Allerdings stehen bisher nur 25 Topics für die Auswertung zur Verfügung, so dass keine weitreichenden Folgerungen erlaubt sind. Es zeigt sich aber, dass die Rolle von Eigennamen im Retrieval vor dem Hintergrund geographischer Eigennamen neu bewertet werden muss. Bisher war bekannt, dass Eigennamen sich positiv auf die Retrieval-Qualität auswirken [Mandl & Womser-Hacker 2005]. Jedoch könnten die Beiträge unterschiedlicher Typen von Eigennamen stark voneinander abweichen.

#### 4 Experimente mit einem geographischen Information Retrieval System

Das System der Informationswissenschaft der Universität Hildesheim hat sich seit mehreren Jahren im Rahmen des Cross Language Evaluation Forum bewährt. Das System baut auf der *open source* Suchmaschine Lucene auf, optimiert dieses durch Mehrwertkomponenten und implementiert die Funktionalität für mehrsprachiges Retrieval [Hackl et al. 2006]. Für die Experimente im Rahmen von geoCLEF wurde es weiter adaptiert.

Ein Systemgestalter eines geographischen Information Retrieval Systems steht derzeit vor zwei Alternativen. Zum einen bietet sich an, traditionelles ad-hoc Retrieval zu betreiben und die geographischen Begriffe als normale Suchterme betrachten. Dies führt nach heutigem Wissensstand oft zu einer guten Leistung. Zum anderen kann ein System geographisches Wissen in starkem Maße formalisieren und darauf Schlussfolgerungen aufbauen.

Der Ansatz der Informationswissenschaft der Universität Hildesheim kombiniert beide Ansätze, wobei jedoch der Schwerpunkt auf Stichwortsuche und Standard-Ansätzen liegt. Im Rahmen eines Standard-Ansatzes werden während der Indexierung geographische Begriffe innerhalb der Dokumente erkannt und in einem gesonderten Index gesammelt. Als klassisches und bewährtes Optimierungsinstrument wurde blindes Relevanz Feedback (BRF) angewandt. Dabei geht das System davon aus, dass die im ersten Suchlauf gefundenen Dokumente relevant sind und extrahiert daraus weitere Suchterme, die zu einer neuen Anfrage führen, welche die endgültigen Ergebnisse liefert. In der Phase des blinden Relevanz Feedbacks nutzt der neue Ansatz nun aber die geographischen Terme sehr spezifisch. Das Relevanz Feedback wirkt sowohl auf den normalen Index als auch auf den geographischen Index, kann jedoch unterschiedlich gewichtet werden (geoBRF). So können durch spezielle Gewichtung von geographischen Termen in den Top-Dokumenten des ersten Durchlaufs mehr geographische Begriffe in die modifizierte Suchanfrage eingehen [Bischoff et al. 2007].

Ein ähnlicher Ansatz zur Optimierung durch flexibles Relevanz Feedback wurde auch erfolgreich in einem System mit kontrolliertem Vokabular eingesetzt, das mehrere Index-Felder umfasste [Berghaus et al. 2006]. Die Experimente basieren auf den Retrievalfunktionen des Lucene-Pakets, dem Lucene-Stemmer für das Deutsche und dem Snowball-Stemmer für das Englische.

Für die Erkennung von Eigennamen wird aufgrund von Evaluierungsergebnissen [Mandl et al. 2005] das maschinelle Lerntool LingPipe eingesetzt, welches anhand eines trainierten Modells Eigennamen identifiziert und in die Kategorien PERSON, LOCATION, ORGANISATION und MISC klassifiziert. Als Modell diente für das Englische das mitgelieferte, an einem englischen Nachrichten-Korpus trainierte News-Modell und für das Deutsche ein selbst entwickeltes Modell.

Die Experimente griffen auf die 25 GeoCLEF Anfragen von 2006 zurück, wobei nur die Felder *title* und *description* genutzt wurden. Eine Testmenge stand nicht zur Verfügung, so dass die gemessenen Werte den Ergebnissen in der Trainingsmenge entsprechen. Tabellen 4 und 5 zeigen die Resultate.

Dokumente für Standard BRF	Terme aus Standard BRF	Dokumente für geoBRF	Terme aus ge- oBRF	MAP
5	100	5	5	0,2082
5	100	15	15	0,2109
5	100	10	10	0,2085
5	100	10	20	0,2092
30	5	15	3	0,1559
10	5	10	3	0,1507
5	10	5	10	0,1718
5	40	0	0	0,1885
5	100	0	0	0,2077

Tabelle 4: Ergebnisse für mono-lingual Englisch

Dokumente für Standard BRF	Terme aus Standard BRF	Dokumente für geoBRF	Terme aus geoBRF	MAP
5	50	15	15	0,2696
10	50	0	0	0,2652
10	10	10	10	0,2807
0	0	0	0	0,256
5	10	10	10	0,2453
15	10	10	10	0,2841
20	10	10	10	0,2625
30	10	10	10	0,2578
5	50	15	5	0,2696
10	50	15	10	0,2652
10	10	0	0	0,2528

Tabelle 5: Ergebnisse für mono-lingual Deutsch

Für beide Sprachen wirkt sich das flexible Relevanz Feedback auf die Retrieval-Qualität aus, wobei die Varianz für das Deutsche geringer ist. Die besten Ergebnisse konnten mit unterschiedlichen Parametersetzungen für das Relevanz Feedback hinsichtlich der Anzahl der berücksichtigten Dokumente und der einbezogenen Terme sowohl für allgemeine als auch geographische Begriffe erzielt werden.

Nachdem in den hier vorliegenden Experimenten geographische Ressourcen mit traditionellen Information Retrieval-Optimierungsverfahren kombiniert wurden, steht für weitere Systemverbesserungen im Rahmen von GeoCLEF 2007 die Integration weiterer Ressourcen im Vordergrund. Unter anderem werden *Wikipedia* zur Erkennung optimaler Übersetzungen von geographischen Eigennamen und die Dienste *Meinestadt.de* und *OpenGeoDB* für geographische Schlussfolgerungen eingesetzt.

## 5 Ausblick: GeoCLEF 2007

GeoCLEF wird auch im Rahmen von CLEF 2007 weitergeführt und umfasst nun zwei Aufgaben. Das in diesem beschriebene ad-hoc Retrieval mit geographischen Einschränkungen ist bisher nicht ausreichend erforscht und der Korpus muss weiter entwickelt werden. Die Aufgabentypen bedürfen aus Evaluierungsperspektive noch einer genaueren Analysen. Die Systementwickler werden noch weitere Ressourcen aufbauen und mit deren Eingliederung in den Information Retrieval Prozess experimentieren.

Des Weiteren führt GeoCLEF 2007 erstmals einen Track zur Klassifizierung von Anfragen ein. Microsoft stellt dafür eine Kollektion von realen Anfragen bereit. Die Teilnehmer sollen die geographischen Anfragen und darin die geographischen Entitäten bestimmen. Ebenso sollen teilnehmende Systeme die gewünschte geographische Beziehung identifizieren (*in, südlich von, entlang, in einer bestimmten Entfernung*). Die Entfernung von Stoppwörtern als Standard-Verarbeitung würde bereits zahlreiche Hinweise auf solche Relationen löschen.

Für weitere Tracks werden Möglichkeiten zum geographischen Retrieval von Photographien analysiert. Dort liegt der geographische Bezug auf der Hand. Bereits im imageCLEF Track vorhandene Kollektionen und Tauschbörsen wie *Flickr* könnten die dafür nötigen Daten liefern. Inhaltsanalyse und Retrieval auf der Basis der Bildunterschriften bzw. Tags sind dann denkbar.

## 6 Literaturverzeichnis

- [Berghaus et al. 2006] Berghaus, Benjamin; Kluck, Michael; Mandl, Thomas (2006): Dynamic Entry Vocabulary Module Implementation and Evaluation. In: Schaaf, Martin; Althoff, Klaus-Dieter (Hrsg): LWA 2006 – Lernen – Wissensentdeckung – Adaptivität Proceedings Workshopwoche GI. 9.-11. Oktober 2006, Universität Hildesheim. Workshop Information Retrieval 2006 [Hildesheimer Informatik-Berichte] S. 94-101.
- [Bischoff et al. 2007] Bischoff, Kerstin; Mandl, Thomas; Womser-Hacker, Christa (2007): Blind Relevance Feedback and Named Entity based Query Expansion for Geographic Retrieval at GeoCLEF 2006. In: Peters, Carol et al. (Eds.): 7<sup>th</sup> Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Alicante, Spain, Revised Selected Papers. Berlin et al.: Springer [Lecture Notes in Computer Science] erscheint. Vorab in: Nardi, Alessandro; Peters, Carol; Vicedo, José Luis (Eds.): CLEF 2006 Working Notes. [http://www.clef-campaign.org/2006/working\\_notes/workingnotes2006/bischoffCLEF2006.pdf](http://www.clef-campaign.org/2006/working_notes/workingnotes2006/bischoffCLEF2006.pdf)
- [Broder 2002] Broder, Andrei (2002): A taxonomy of web search. In: ACM SIGIR Forum vol. 36(2) S. 3 – 10.

- [Buckley & Voorhees 2005] Buckley, Chris; Voorhees, Ellen (2005). Retrieval System Evaluation. In: TREC: Experiment and Evaluation in Information Retrieval. Cambridge & London: MIT Press. S. 53-75.
- [Gey et al. 2007] Gey, Fredric; Larson, Ray; Sanderson, Mark; Bishoff, Kerstin; Mandl, Thomas; Womser-Hacker, Christa; Santos, Diana; Rocha, Paulo; Di Nunzio, Giorgio; Ferro, Nicola (2007): GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. In: Peters, Carol et al. (Eds.): 7<sup>th</sup> Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Alicante, Spain, Revised Selected Papers. Berlin et al.: Springer [Lecture Notes in Computer Science] erscheint. Vorab in: Nardi, Alessandro; Peters, Carol; Vicedo, José Luis (Eds.): CLEF 2006 Working Notes. [http://www.clef-campaign.org/2006/working\\_notes/workingnotes2006/geyOCLEF2006.pdf](http://www.clef-campaign.org/2006/working_notes/workingnotes2006/geyOCLEF2006.pdf)
- [Grivolla & Mori 2005] Grivolla, J.; Jourlin, P. and de Mori, R. (2005): Automatic Classification of Queries by Expected Retrieval Performance. In: Predicting Query Difficulty – Methods and Applications. Workshop at the 28<sup>th</sup> Annual International ACM Conference on Research and Development in Information Retrieval (SIGIR) Salvador, Bahia, Brazil, August 19, 2005. <http://www.haifa.il.ibm.com/sigiro5-qp/>
- [Hackl et al. 2006] Hackl, René; Mandl, Thomas; Womser-Hacker, Christa (2006): Ad-hoc Mono- and Multilingual Retrieval Experiments at the University of Hildesheim. In: Peters, Carol; Gey, Fredric C.; Gonzalo, Julio; Jones, Gareth J.F.; Kluck, Michael; Magnini, Bernardo; Müller, Henning; Rijke, Maarten de (Eds.). Accessing Multilingual Information Repositories: 6<sup>th</sup> Workshop of the Cross-Language Evaluation Forum, CLEF 2005, Vienna, Austria, Revised Selected Papers. Berlin et al.: Springer [Lecture Notes in Computer Science 4022] S. 37-43.
- [Mandl 2006] Mandl, Thomas (2006): Neue Entwicklungen bei den Evaluierungsinitiativen im Information Retrieval. In: Mandl, Thomas; Womser-Hacker, Christa (Hrsg.): Effektive Information Retrieval Verfahren in der Praxis: Proceedings Vierter Hildesheimer Evaluierungs- und Retrievalworkshop (HIER 2005) Hildesheim, 20.7.2005. Konstanz: Universitätsverlag [Schriften zur Informationswissenschaft 45] S. 117-128.
- [Mandl & Womser-Hacker 2005] Mandl, Thomas; Womser-Hacker, Christa (2005): The Effect of Named Entities on Effectiveness in Cross-Language Information Retrieval Evaluation. In: Applied Computing 2005: Proceedings of 2005 ACM SAC Symposium on Applied Computing (SAC). Information Access and Retrieval (IAR) Track. Santa Fe, New Mexico, USA. March 13.-17. 2005. ACM Press. S. 1059-1064.
- [Mandl et al. 2005] Mandl, Thomas; Schneider, René; Schnetzler, Pia; Womser-Hacker, Christa (2005): Evaluierung von Systemen für die Eigennamenerkennung im cross-lingualen Information Retrieval. In: Fisseni, Bernhard; Schmitz, Hans-Christian; Schröder, Bernhard; Wagner, Petra (Hrsg.): Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen: Beiträge der zur GLDV Tagung 2005 in Bonn (Gesellschaft für linguistische Datenverarbeitung) 30.3.-01.04. [Sprache, Sprechen und Computer/Computer Studies in Language and Speech 8] Frankfurt a. M. et al.: Peter-Lang. S. 145-157.

- [Mandl & Womser-Hacker 2006] Mandl, Thomas; Womser-Hacker, Christa (2006):  
Information Retrieval – eine Schlüsseltechnologie in der Wissensgesellschaft. In: WISU:  
Das Wirtschaftsstudium. Erscheint.
- [Peters et al. 2006] Peters, Carol; Gey, Fredric; Gonzalo, Julio; Jones, Gareth J.F.; Kluck,  
Michael; Magnini, Bernardo; Müller, Henning; Rijke, Maarten de (2006) (Eds.).  
Accessing Multilingual Information Repositories: 6<sup>th</sup> Workshop of the Cross-Language  
Evaluation Forum, CLEF 2005, Vienna, Austria, Revised Selected Papers. Berlin et al.:  
Springer [Lecture Notes in Computer Science 4022]

# Personalisierung mittels informationsbewertender Profile\*

*Tobias Galliat*

Institut für Informationswissenschaft,  
Fachhochschule Köln,  
Claudiusstr. 1, 50678 Köln  
*tobias.galliat@fh-koeln.de*

## **Zusammenfassung**

Die automatisierte Personalisierung von Web-Angeboten erfordert eine schnelle Profilbildung. Der Open Profiling Standard versucht das Kaltstart-Problem zu vermeiden, indem Nutzer-Profile normiert werden, so dass sie zwischen unterschiedlichen Profilingssystemen übertragen werden können. In Fällen wo dies nicht praktikabel ist, müssen zur anfänglichen Profilbildung auch Informationen niedrigen Qualitätsgrades – beispielsweise Prognosen – verwendet werden. Dies setzt jedoch ein Profilmodell voraus, bei dem der Informationsgrad geeignet bewertet wird. Hierzu wird ein informations-theoretisch motiviertes Konzept vorgestellt, das zudem eine beschleunigte Profilbildung durch Wechselwirkung zwischen Profilen vorsieht.

## **I Einleitung**

Die Attraktivität des Internets hängt offensichtlich eng damit zusammen, dass es prinzipiell für alle Bedürfnisse von Menschen in Bezug auf den Erwerb und den Austausch von Information, Waren und Dienstleistungen entsprechende Angebote bereithält. Allerdings ist in der Regel ein erheblicher, nicht immer von Erfolg gekrönter Suchaufwand erforderlich, bevor der Internet-Nutzer auf die für ihn relevanten Angebote und Inhalte zugreifen kann. Ob und in welchem Umfang die Suche erfolgreich ist, hängt dabei stark von seinen individuellen Fähigkeiten im Umgang mit dem Internet sowie der Zeit ab, die er bereit ist in die Suche zu investieren.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 27-41.

Es ist deshalb wenig verwunderlich, dass neben der Bereitstellung von effizienten Suchmaschinen die Personalisierung von Internet-Angeboten mittels Profilen – zum Beispiel via Portal-Seiten – als Ausweg aus diesem Dilemma angesehen wird. Zu unterscheiden ist dabei zwischen einer manuellen und einer automatisierten Personalisierung<sup>1</sup>. Die manuelle Personalisierung, die das Festlegen von bevorzugten Web-Seiten, aber auch die detaillierte Auswahl von Angeboten auf einer Portal-Seite umfassen kann, nimmt der Nutzer bewusst und unmittelbar selber vor. Dies setzt voraus, dass er seine Bedürfnisse genau kennt und den entsprechenden Angeboten zuordnen kann. Offensichtlich verkürzt die manuelle Personalisierung hauptsächlich die Zeit der Suche nach im Wesentlichen bekannten Angeboten und bereitet diese in strukturierter Form auf. Für den weniger erfahrenen Nutzer bietet sie hingegen keine wirkliche Lösung seines Suchproblems. Er ist sich seiner Bedürfnisse teilweise gar nicht bewusst bzw. kann diesen keine oder nur wenige relevante Angebote unmittelbar zuordnen. Bei der automatisierten Personalisierung mittels Profilen ist dies auch nicht notwendig, denn hier wirkt der Nutzer nur mittelbar, explizit oder implizit mit, indem er beispielsweise durch Angabe seines Geburtsdatums, die Bewertung von Angeboten oder sein aufgezeichnetes Klickverhalten, die Ausprägungen vorgegebener Attribute – wie z. B. das Alter, die Interessen oder die mittlere Verweildauer im Internet – festlegt. Hieraus lässt sich ein Nutzer-Profil ableiten, dem dann automatisch Angebote und Inhalte zugeordnet werden, die in der Vergangenheit für Nutzer mit ähnlichen Profilen – die über das so genannte „Profil-Matching“ ermittelt werden – relevant gewesen sind. Dieser Vorgang wird in der Literatur üblicherweise als „kollaboratives Filtern“ [Breese et al. 98], die entsprechende Personalisierungs-Software als „Recommender Systeme“ [Terveen et al. 01] bezeichnet<sup>2</sup>.

Nun muss es das Ziel jedes Anbieters von personalisierten Web-Services sein, das so genannte Kaltstartproblem zu lösen, d. h. eine möglichst schnelle Profilbildung bei jedem neuen Nutzer – optimalerweise nach dem ersten Klick – zu erreichen und danach die Profile der wiederkehrenden Nutzer – durch das so genannte „Profil-Influencing“ – immer weiter zu verfeinern und aktuell zu halten. Mit dem „Open Profiling Standard (OPS)“ hat es schon frühzeitig Versuche gegeben eine Normierung der Profile zu schaffen [Dunn et al. 97], so dass Web-Anbieter ein bereits be-

---

<sup>1</sup> In der Literatur zum Thema „Benutzermodellierung“ werden alternativ oft die – im Wesentlichen jeweils inhaltlich deckungsgleichen – Begriffe „(benutzer)-adaptierbar“ und „(benutzer)-adaptiv“ verwendet [Kobsa 04].

<sup>2</sup> Es gibt allerdings auch Recommender Systeme, die mit profilunabhängigen Filterverfahren, wie z. B. eigenschaftsbasierten Filtern arbeiten [van Meteren et al. 00].



stehendes Profil – bei entsprechender Freigabe durch den Nutzer – zur Personalisierung ihres Angebotes verwenden können, ohne dass der Nutzer die Anbieter-Seite bereits früher besucht bzw. dort einen umfangreichen Prozess der Profilbildung absolviert hat. In der Praxis hat der OPS-Ansatz mit diversen Schwierigkeiten zu kämpfen, wie beispielsweise der zunehmenden Sensibilität der Nutzer bezüglich der Weitergabe ihrer persönlichen Daten, der in der Normierung begründet liegenden Beschränktheit der Profilbildung oder der Aktualität der Profile. Damit bietet der OPS-Ansatz in vielen Fällen keine wirklich befriedigende Lösung für das Kaltstartproblem von Profilingssystemen.

Im Folgenden soll deshalb das Konzept einer automatisierten Personalisierung mittels informationsbewertender Profile vorgestellt werden, bei dem die Profilbildung dadurch beschleunigt wird, dass zum einen nicht nur den Nutzern, sondern auch den Informations- oder Warenangeboten – den so genannten „Items“ – Profile zugeordnet werden, die sich wechselseitig beeinflussen können<sup>3</sup>. Dadurch wird die Trennung zwischen der Profilbildung und dem Prozess des kollaborativen Filterns aufgehoben und im Gegensatz zu vielen herkömmlichen Ansätzen, besteht die Möglichkeit neue Angebote sofort in das System zu integrieren<sup>4</sup>. Zum anderen sieht das Konzept vor, dass jede zur Profilbildung verwendete Attribut-Ausprägung mit einem so genannten „Informationsgrad“ versehen wird, der bewertet wie zuverlässig die Festlegung der Ausprägung ist<sup>5</sup>. Um die Interpretation zu erleichtern, bietet es sich an, den Wertebereich für den Informationsgrad auf das Intervall  $[0,1]$  festzusetzen. Eine Attribut-Ausprägung erhält dann einen Informationsgrad nahe bei 0, wenn die Aussage, dass die Attribut-Ausprägung dem Profil zugeordnet werden kann, keinen oder nur eine sehr geringe Validität besitzt. Ein Wert nahe bei 1 ist hingegen als eine (fast) sichere Information zu interpretieren. Der Informationsgrad ist dabei nicht unveränderlich über die Zeit. Beispielsweise lässt sich eine Adressangabe durch den erfolgreichen Versand eines bestellten Produktes verifizieren oder das Nutzer-Interesse an einem bestimmten Angebot durch dessen regelmäßige Besuche der entsprechenden Web-Seiten bestätigen. Damit wird es möglich, instantan Profile auf der Basis von Informationen niedrigen Qualitätsgrades – beispielsweise Hypothesen – zu bilden und sie dann im Laufe der Zeit zu verfeinern und zu validieren.

---

<sup>3</sup> In der neueren Literatur um Thema „Collaborative Filtering“ finden sich zunehmend Ansätze zur Verbindung von Nutzer- und Item-Profilen [Wang et al. 06].

<sup>4</sup> Es existieren auch Verfahren, die dies auf anderem Wege erreichen [Schein et al. 02].

<sup>5</sup> In anderen Ansätze wird die Nutzer-Zuverlässigkeit bewertet [Massa et al. 04].

Basis des entwickelten Konzeptes ist ein Modell für Profilingssysteme, bei dem – in Analogie zum aus dem Information Retrieval bekannten Vektorraummodell – Profile durch Objekte in einem mehrdimensionalen Raum – im Folgenden als Profil-Raum bezeichnet – repräsentiert werden, wobei jedem Attribut mit seiner endlichen Anzahl von zulässigen Ausprägungen jeweils eine Dimension des Raumes zugeordnet ist<sup>6</sup>. Neu ist hier die Modellierung des Informationsgrades als so genannte „charakteristische Dimension“. In diesem Kontext lässt sich dann das Profil-Influencing als die regelbasierte Veränderung der Objekte des Profil-Raumes über die Zeit und durch Wechselwirkung untereinander interpretieren. Das Profil-Matching wiederum lässt sich als die Bestimmung von im Profil-Raum nahe beieinander liegenden Objekten auffassen. Dabei wird der Grad der „Nähe“ mittels eines Distanzmaßes bestimmt, das den Informationsgrad sinnvoll berücksichtigt und auch der Tatsache gerecht wird, dass die Mehrzahl der Objekte im Profilraum für viele Dimensionen keine echte Ausprägung besitzen bzw. vielen Ausprägungen ein Informationsgrad nahe bei 0 zugeordnet wird. Diese Vorgaben führen zur Verwendung einer bereits früher entwickelten informationsbewertenden Distanz [Galliat 99] im Zusammenspiel mit einer Adaption des Ähnlichkeits-Maßes von Tanimoto.

Die Präsentation des Konzeptes erfolgt in drei Schritten: In einem ersten Schritt wird durch Formulierung der gewünschten Eigenschaften eines Profils ein geeigneter Profil-Raum konstruiert, dann werden die Regeln für das Profil-Influencing beschrieben und im letzten Schritt ein zum Profil-Matching geeignetes informationsbewertendes Distanzmaß vorgestellt. Abschließend erfolgt noch ein Ausblick auf mögliche Erweiterungen des Konzeptes im Hinblick auf eine praktische Realisierung.

## 2 Konstruktion des mehrdimensionalen Profil-Raumes

Bei der Konstruktion des Profil-Raumes, sind folgende gewünschte Eigenschaften eines Profils zu berücksichtigen:

- Die Ausprägungen der Attribute besitzen nicht zwingend eine natürliche Ordnung, d. h. die Dimensionen des Profil-Raumes sind im Allgemeinen weder ordinal noch metrisch skaliert.

---

<sup>6</sup> Vorausgesetzt wird dabei zunächst, dass es zu jedem Attribut eine „Missing“-Ausprägung gibt, d. h. einen Wert der gesetzt wird, wenn keine der echten Ausprägungen zutrifft. Durch die Berücksichtigung des Informationsgrades wird diese künstliche Ausprägung überflüssig.

- Ein Profil kann mehrere Ausprägungen eines Attributes besitzen<sup>7</sup>, wird also nicht durch einen Punkt, sondern durch ein komplexeres Objekt im Profil-Raum repräsentiert.
- Damit einem Profil eine Attribut-Ausprägung zugeordnet werden kann, ist ein rechtfertigendes Ereignis – der so genannte „Hit“ – notwendig. Je nach dem ob dieses Ereignis beobachtet<sup>8</sup>, aufgrund einer gezielten Befragung ermittelt<sup>9</sup> oder durch eine Prognose – oft auf der Basis von soziodemographischen Daten – ausgelöst wird<sup>10</sup>, spricht man von einem Beobachtungsereignis, einem Befragungsereignis oder einem Prognoseereignis. Prognoseereignisse sind dabei in der Regel die Folge von Beobachtungs- oder Befragungsereignissen. Deshalb bestehen hier auch oft Abhängigkeiten zwischen den zugehörigen Attribut-Ausprägungen.
- Die Aussage, dass einem Profil eine bestimmte Attribut-Ausprägung zugeordnet werden kann, ist oft mit Unsicherheit behaftet. Deshalb soll für jede Attribut-Ausprägung die einem Profil zugeordnet wird, die Zuverlässigkeit der Information bewertet und als Informationsgrad vermerkt sein, d. h. jeder Attribut-Ausprägung wird eine zusätzliche charakteristische Dimension zugeordnet, die den Informationsgrad widerspiegelt.
- Die Attribut-Ausprägungen und die ihnen zugeordneten Informationsgrade können sich im Laufe der Zeit ändern<sup>11</sup>. Die Objekte im Profil-Raum sind also nicht statisch, sondern dynamisch. Aus diesem Grund ist es zweckmäßig den Attribut-Ausprägungen weitere charakteristische Dimensionen zuzuordnen. So ist es sinnvoll die Kategorie des rechtfertigenden Ereignisses als „HitCategory“, den Zeitpunkt des letzten rechtfertigenden Ereignisses als „HitTime“ und die durchschnittliche Dauer zwischen zwei rechtfertigenden Ereignissen als „AverageHitToHitTime“ zu vermerken<sup>12</sup>.

---

<sup>7</sup> Ein Nutzer interessiert sich für Reisen nach Frankreich und Schweden. Seinem Profil müssen also für das Attribut „bevorzugte Reiseziele“ mindestens zwei gültige Ausprägungen zugeordnet werden.

<sup>8</sup> Z. B. durch das Anklicken eines Werbebanners.

<sup>9</sup> Z. B. im Rahmen des einmaligen Anmeldeprozesses für einen Web-Service

<sup>10</sup> Beispielsweise lässt sich die Ausprägung für das Attribut „Kreditwürdigkeit“ aus einem Abgleich von Personendaten mit der Datenbank einer Kredit-Auskunftei ermitteln.

<sup>11</sup> Eltern interessieren sich z. B. in der Regel nur für einen gewissen Zeitraum für Babysachen. Sobald die Kinder älter sind, werden entsprechende Attribut-Ausprägungen für ihre Profile keine oder nur noch eine sehr geringe Gültigkeit besitzen.

<sup>12</sup> Üblicherweise gilt: je häufiger ein rechtfertigendes Ereignis beobachtet wird, umso zuverlässiger ist die Zuordnung zu einem Profil. Allerdings sollte berücksichtigt werden, wie häufig ein derartiges rechtfertigendes Ereignis überhaupt zu erwarten ist.

Aus den aufgeführten Eigenschaften lässt sich folgendes Modell für den Profil-Raum ableiten:

Sei  $A_1, \dots, A_k$  die Menge der Attribute, die zur Profilbildung verwendet werden sollen und für jedes Attribut  $A_i$  bezeichne  $M_i := M(A_i) := \{a_{i1}, \dots, a_{in(i)}\}$  die Menge der zulässigen Ausprägungen. Diese Attribute bilden die Attribut-Dimensionen des Profil-Raumes und die Mengen  $M_i$  die zugehörigen Wertebereiche. Zusätzlich existieren für jede Attribut-Ausprägung  $a_{ij}$  charakteristische Dimensionen für den Informationsgrad, die Kategorie und den Zeitpunkt des letzten rechtfertigenden Ereignisses sowie die durchschnittliche Dauer zwischen zwei rechtfertigenden Ereignissen.

Besitzt ein Profil mehr als eine Ausprägung für ein Merkmal lässt es sich offensichtlich nicht als Punkt im Profil-Raum interpretieren. Wir modellieren deshalb ein Profil als komplexes Objekt des Profil-Raumes, dem für jede (Attribut-)Dimension eine Menge von Ausprägungen zugeordnet werden können.

Wie nun diese Objekte – ausgelöst durch rechtfertigende Ereignisse – konkret gebildet werden und sich über die Zeit verändern bzw. sich wechselseitig beeinflussen können, wird im nächsten Schritt erläutert.

### 3 Regeln für das Profil-Influencing

Beim Profil-Influencing lassen sich die folgenden Situationen unterscheiden:

- Erstmalige Zuordnung von Attribut-Ausprägungen zu einem Profil, sowie die Festlegung der zugehörigen charakteristischen Werte, aufgrund eines rechtfertigenden Ereignisses.
- Änderung der charakteristischen Werte von bereits zugeordneten Attribut-Ausprägungen durch das Ausbleiben bzw. das Eintreten von weiteren rechtfertigenden Ereignissen.
- Wechselwirkung zwischen Profilen, die zur Zuordnung von Attribut-Ausprägungen und Änderung von charakteristischen Werten führt.

#### 3.1 Erstmalige Zuordnung von Attribut-Ausprägungen

Betrachten wir zunächst die Situation einer erstmaligen Zuordnung von Attribut-Ausprägungen.

Zu einem Zeitpunkt  $T$  erfolge ein rechtfertigendes Ereignis für die Zuordnung der o.b.d.A eindeutigen Ausprägung  $x$  für das Attribut  $A$  zum Profil  $P$ . Sei  $x^*(P)$  der Informationsgrad mit dem  $x$  dem Profil  $P$  zugeordnet werden kann. Mit  $HC(x, P)$ ,

$HT(x,P)$  und  $\emptyset H_2 HT(x,P)$  seien ferner die HitCategory, die HitTime und die AverageHitToHitTime von  $x$  bezüglich des Profils  $P$  bezeichnet.

Zusätzlich sei  $GHT(x)$  bzw.  $G\emptyset H_2 HT(x)$  als globale durchschnittliche HitTime bzw. globale durchschnittliche AverageHitToHitTime der Ausprägung  $x$  über alle bereits im System befindlichen Profile definiert<sup>13</sup>:

$$(1) \quad GHT(x) := \bigcap_{HT(x,P) \text{ definiert}} HT(x,P) \quad \text{bzw.}$$

$$(2) \quad G\emptyset H_2 HT(x) := \bigcap_{\emptyset H_2 HT(x,P) \text{ definiert}} \emptyset H_2 HT(x,P) .$$

Als erstes wird die Kategorie  $HC(x,P)$  des rechtfertigenden Ereignisses für  $x$  bezüglich  $P$  vermerkt. Handelt es sich hierbei um ein Beobachtungs- oder Befragungsergebnis, setzt man als nächstes den Informationsgrad auf einen a priori festgelegten Wert. Wie die späteren Ausführungen zur Erhöhung des Informationsgrades bei der Beobachtung weiterer rechtfertigender Ereignisse zeigen werden, bietet sich hier der Wert 0,5 an. Durch Auswertung externer Quellen bzw. die Anwendung von Validierungsregeln<sup>14</sup> kann der Informationsgrad ggf. noch herauf oder herabgesetzt werden. Als HitTime  $HT(x,P)$ , wird der aktuelle Zeitpunkt  $T$  vermerkt. Für  $\emptyset H_2 HT(x,P)$  setzt man den Wert  $G\emptyset H_2 HT(x)$  ein<sup>15</sup>. Handelt es sich hingegen um ein Prognoseereignis, geht man anders vor. Derartige Ereignisse sind immer mit früher stattgefundenen Ereignissen verknüpft, d. h. die Zuordnung einer Ausprägung  $x$  zum Profil  $P$  erfolgt aufgrund anderer Ausprägungen, die  $P$  bereits früher zugeordnet worden sind. Bezeichne deshalb im Folgenden  $D(x,P)$  die Menge der Attribut-Ausprägungen von denen  $x$  abhängt<sup>16</sup>.

Die charakteristischen Werte von  $x$  bezüglich  $P$  sind natürlich ebenfalls abhängig von den entsprechenden Werten der Ausprägungen in  $D(x,P)$ . In einem konservativen Ansatz wählen wir jeweils den minimalen Wert:

$$(3) \quad x^*(P) := \min_{\tilde{x} \in D(x,P)} \tilde{x}^*(P) ,$$

$$(4) \quad HT(x,P) := \min_{\tilde{x} \in D(x,P)} HT(\tilde{x}, P) ,$$

$$(5) \quad \emptyset H_2 HT(x,P) := \min_{\tilde{x} \in D(x,P)} \emptyset H_2 HT(\tilde{x}, P) .$$

<sup>13</sup> Das  $\cap$ -Symbol ist hier wie folgt zu interpretieren:

$$\bigcap_{\lambda \in I} f(\lambda) := \sum_{i=1}^r \frac{1}{r} f(\lambda_i) \quad \text{mit } I = \{\lambda_1, \dots, \lambda_r\}.$$

<sup>14</sup> Z. B. empirischer Zusammenhang zwischen Vornamen und Geburtsjahrgang.

<sup>15</sup> Der Wert von  $G\emptyset H_2 HT(x_i)$  darf anfangs undefiniert bzw. eine Konstante sein.

<sup>16</sup> Beispielsweise kann die Ausprägung für das Attribut „Kaufkraft“ von den Ausprägungen der Attribute „Wohnort“ und „Alter“ abhängen.

Als nächstes stellt sich nun die Frage, welche Auswirkungen das Auftreten bzw. das Ausbleiben von rechtfertigenden Folgeereignissen für  $x$  bezüglich  $P$  auf die zugehörigen charakteristischen Werte haben.

### 3.2 Regelmäßige Aktualisierung von charakteristischen Werten

Beobachtete Ereignisse die eine Aufnahme von  $x$  ins Profil  $P$  rechtfertigen sind im Allgemeinen nicht singulär, sondern können im Laufe der Zeit immer wieder auftreten. Vielmehr sollten sie sogar regelmäßig auftreten, denn sonst muss von einer abnehmenden Gültigkeit der Attribut-Ausprägung für das betrachtete Profil ausgegangen werden. Die Häufigkeit des Auftretens in einem bestimmten Zeitraum sollte dabei nicht absolut bewertet werden<sup>17</sup>, sondern relativ zu der in der Vergangenheit beobachteten Frequenz und der Häufigkeit, mit der entsprechende Beobachtungen bei anderen Profilen erfolgen. Der Informationsgrad einer Attribut-Ausprägung für ein Profil hängt also in geeigneter Weise vom Zeitpunkt des letzten rechtfertigenden Ereignisses sowie der in der Vergangenheit beobachteten durchschnittlichen Zeit zwischen zwei rechtfertigenden Ereignissen ab, und zwar sowohl bezogen auf das gegebene Profil als auch über alle Profile hinweg.

Befragungsereignisse sind hingegen tendenziell singulärer Natur<sup>18</sup>. Hier bleibt in der Folge der Informationsgrad über die Zeit konstant, es sei denn es gibt konkrete Hinweise, die eine Änderung begründen.

Anders sieht die Situation bei Attribut-Ausprägungen  $x$  aus, die aufgrund eines Prognoseereignisses dem Profil  $P$  zugeordnet worden sind. Solange sich die Kategorie von  $x$  nicht ändert, müssen die charakteristischen Werte bei jeder regelmäßigen Aktualisierung von Attribut-Ausprägungen aus der Menge  $D(x,P)$  entsprechend den Regeln für die erstmalige Zuordnung<sup>19</sup> ebenfalls aktualisiert werden.

Im Folgenden sei nun  $x$  eine Attribut-Ausprägung für die bezüglich des Profils  $P$  letztmalig zum Zeitpunkt  $HT(x,P)$  ein rechtfertigendes Ereignis beobachtet worden ist. Ferner sei  $T$  der aktuelle Beobachtungszeitpunkt.

---

<sup>17</sup> Nutzer A besucht die Web-Seiten eines E-Commerce-Anbieters täglich, Nutzer B jedoch nur einmal wöchentlich. Die Häufigkeit von rechtfertigenden Ereignissen wird somit bei Nutzer A größer sein als bei Nutzer B. Daraus lässt sich aber keine entsprechende Aussage über die Gültigkeit der Attribut-Ausprägungen treffen, die den Profilen zugeordnet sind.

<sup>18</sup> Die wiederholte Befragung führt bei Kunden oft zu Irritationen und wird von diesen in der Regel abgelehnt. Sollte dennoch erfolgreich eine erneute Befragung durchgeführt worden sein, so werden einfach die alten durch die neuen Attribut-Ausprägungen ersetzt.

<sup>19</sup> Siehe Formeln (3) – (5).

### 3.2.1 Kein rechtfertigendes Ereignis zum Zeitpunkt $T$

Wir betrachten zunächst den Fall, dass zum Zeitpunkt  $T$  kein rechtfertigendes Ereignis für  $x$  bezüglich  $P$  beobachtet wird. Wenn nun der Zeitraum zwischen  $HT(x,P)$  und  $T$  größer ist als die AverageHitToHitTime von  $x$  bezüglich  $P$ , d. h. wenn gilt:  $\emptyset H_2 HT(x,P) \leq T - HT(x,P)$ , dann reduzieren wir den Informationsgrad („regelmäßiger Downgrade“):

$$(6) \quad x^*(P) := x^*(P) - (0,5 / (2t - \emptyset H_2 HT(x,P))) x^*(P).$$

Dabei ist  $t$  das Maximum aus  $\emptyset H_2 HT(x,P)$  und dem halben globalen durchschnittlichen Zeitraum zwischen zwei rechtfertigenden Ereignissen für  $x$ , d. h.  $t := \max(\emptyset H_2 HT(x,P), G\emptyset H_2 HT(x) / 2)$ .

Außer für Profile, bei denen sehr selten rechtfertigende Ereignisse beobachtet werden, gilt  $t = \emptyset H_2 HT(x,P)$ , d. h. der Informationsgrad wird um den Faktor  $0,5/\emptyset H_2 HT(x,P)$  reduziert<sup>20</sup>. Je regelmäßiger also in der Vergangenheit ein rechtfertigendes Ereignis für  $x$  bezüglich  $P$  zu beobachten war, umso stärker ist der Downgrade, wenn es im weiteren Verlauf ausbleibt.

Sinkt der Informationsgrad von  $x$  bezüglich  $P$  nach dem Downgrade auf einen Wert nahe bei 0, d. h. gilt  $x^*(P) < \varepsilon$  für sehr kleines  $\varepsilon > 0$ , wird die Ausprägung automatisch aus dem Profil entfernt und  $HT(x,P)$ , sowie  $\emptyset H_2 HT(x,P)$  als undefiniert gekennzeichnet.

Wenn der nächste erwartete Zeitpunkt für ein rechtfertigendes Ereignis hingegen noch nicht erreicht ist, wenn also gilt  $\emptyset H_2 HT(x,P) > T - HT(x,P)$ , dann bleibt  $x^*(P)$  unverändert<sup>21</sup>.

Die charakteristischen Werte  $HT(x,P)$  und  $\emptyset H_2 HT(x,P)$  ändern sich nicht.

### 3.2.2 Rechtfertigendes Ereignis zum Zeitpunkt $T$

Betrachten wir nun die Situation, dass zum Zeitpunkt  $T$  ein rechtfertigendes Ereignis für  $x$  bezüglich  $P$  beobachtet wurde. In diesem Fall setzen wir:

$$(7) \quad x^*(P) := x^*(P) + (1 - x^*(P)) / 2.$$

<sup>20</sup> Sinnvollerweise arbeitet man mit einem diskreten Zeitgeber, d. h.  $T$  wird immer um eine Zeiteinheit erhöht. Es gilt also  $\emptyset H_2 HT(x,P) \geq 1$  und damit nach dem regelmäßigen Downgrade auch immer  $x^*(P) \geq 0$ .

<sup>21</sup> Dies ist insbesondere auch dann der Fall, wenn  $\emptyset H_2 HT(x,P)$  noch undefiniert ist.

Der Informationsgrad wird also um den Wert  $(1-x(P))/2$  erhöht<sup>22</sup>. Zusätzlich müssen auch noch die Werte  $HT(x,P)$  und  $\emptyset H_2 HT(x,P)$  angepasst werden.

Falls  $\emptyset H_2 HT(x,P)$  undefiniert ist, setzen wir

$$(8) \quad \emptyset H_2 HT(x,P) := T - HT(x,P),$$

andernfalls wird ein neuer Durchschnittswert berechnet<sup>23</sup>:

$$(9) \quad \emptyset H_2 HT(x,P) := (\emptyset H_2 HT(x,P) + T - HT(x,P)) / 2.$$

Im Anschluss aktualisieren wir dann noch die HitTime auf den aktuellen Zeitpunkt, d. h.  $HT(x,P) := T$ .

### 3.3 Wechselwirkung zwischen Profilen

Der Aufbau von Profilen durch die bisher beschriebenen Regeln ist sehr zeitaufwendig, da sehr viele rechtfertigende Ereignisse benötigt werden. Durch den Einsatz von Prognosen, lässt sich dieser Prozess schon deutlich beschleunigen. Allerdings ist die Formulierung von Prognosen nicht trivial, sondern es müssen geeigneten Datenquellen zur Verfügung stehen und anspruchsvolle Techniken der statistischen Datenanalyse und des Data-Mining, wie z. B. Cluster-Verfahren [Ungar et al. 98] genutzt werden. Des Weiteren sind Attribut-Ausprägungen, die aufgrund von Prognosen einem Profil zugeordnet werden, auch immer nur so valide, wie es die der Prognose zugrunde liegenden Attribut-Ausprägungen sind. Können einem Profil am Anfang also nur wenige durch Beobachtung oder Befragung ermittelte Attribut-Ausprägungen zugeordnet werden – wie dies üblicherweise oft der Fall ist – so lassen sich auch nur schwer Prognosen mit hohem Informationsgrad abgeben.

Einen Ausweg aus diesem Dilemma bietet die Wechselwirkung zwischen Profilen. Hierzu ist es erforderlich nicht nur Nutzer-, sondern auch Item-Profile anzulegen, die sich dann wechselseitig beeinflussen können<sup>24</sup>. Dadurch erfolgt ein schneller Aufbau von Profilen sowohl für neue Nutzer, als auch für neue Items. Bei der Mo-

---

<sup>22</sup> Offensichtlich wird so sichergestellt, dass stets gilt:  $x(P) \leq 1$ . Für den Fall, dass  $x$  dem Profil  $P$  zuvor nicht zugeordnet werden konnte, also  $x(P)=0$  gilt, erhält man nach dem rechtfertigenden Ereignis den Wert  $x(P)=0,5$ . Dies ist konsistent mit der vorgeschlagenen Bewertung bei erstmaliger Beobachtung eines rechtfertigenden Ereignisses für  $x$  bezüglich  $P$ .

<sup>23</sup> Es handelt sich um keine echte Durchschnittsbildung, sondern der letzte Zeitraum wird immer stark übergewichtet. Damit wird eine schnellere Änderung der charakteristischen Werte ermöglicht und auch die Datenhaltung vereinfacht.

<sup>24</sup> Web-Links können z. B. mit Profilen versehen werden. Wenn nun der Nutzer auf einen solchen Link klickt, beeinflussen sich sein Profil und das Link-Profil wechselseitig.



dellierung gibt es keinen Unterschied zwischen einem Nutzer- und einem Item-Profil, beides sind Objekte im Profil-Raum. Wir formulieren deshalb als nächstes Regeln für die Wechselwirkung zwischen Objekten P und Q des Profil-Raumes. Dabei betrachten wir jede Attribut-Dimension separat.

Profil P besitze für Attribut A die Ausprägungen  $x_1, \dots, x_s$  und Profil Q die Ausprägungen  $y_1, \dots, y_r$  mit jeweils positivem Informationsgrad. Alle Attribut-Ausprägungen in Q die nicht in P liegen, werden auch in P eingefügt, allerdings mit Informationsgrad 0. Entsprechend verfährt man mit Attribut-Ausprägungen in P die nicht in Q liegen. Anschließend sortiert man die Ausprägungen so um, dass  $x_j = y_j$  für  $j = 1, \dots, n$ , wobei n die Anzahl der gemeinsamen Attribut-Ausprägungen in Profil P und Q bezeichne.

Die wechselseitige Beeinflussung von P und Q erfolgt nun über eine Veränderung des Informationsgrades. Dies ist gerechtfertigt, da es sich bei der Wechselwirkung ja offensichtlich um die Folge eines beobachteten Ereignisses handelt. Für  $j = 1, \dots, n$  setzt man:

$$(10) \quad x_j^*(P) := x_j^*(P) + \Delta(y_j^*(Q), x_j^*(P))$$

$$\text{mit } \Delta(y_j^*(Q), x_j^*(P)) := (y_j^*(Q) - x_j^*(P)) ((1 - d_j^*(P, Q)) + (1 - x_j^*(P))) / 2,$$

$$(11) \quad y_j^*(Q) := y_j^*(Q) + \Delta(x_j^*(P), y_j^*(Q))$$

$$\text{mit } \Delta(x_j^*(P), y_j^*(Q)) := (x_j^*(P) - y_j^*(Q)) ((1 - d_j^*(Q, P)) + (1 - y_j^*(Q))) / 2.$$

Dabei bezeichnet  $d_j^*(P, Q)$  die informationsbewertende Distanz [Galliat 99] zwischen den Profilen P und Q für die Ausprägung  $x_j (= y_j)$  entlang der betrachteten (Attribut-)Dimension. Wir definieren:

$$(12) \quad d_j^*(P, Q) := \log_2(2 - (x_j^*(P) y_j^*(Q))^{1/2})$$

Offensichtlich gilt  $d_j^*(P, Q) \in [0, 1]$  sowie  $d_j^*(P, Q) = 0$  gdw.  $x_j^*(P) y_j^*(Q) = 1$  und  $d_j^*(P, Q) = 1$  gdw.  $x_j^*(P) y_j^*(Q) = 0$ .

Man verifiziert leicht, dass keine Änderung des Informationsgrades von  $x_j$  bezüglich P erfolgt, wenn  $y_j^*(Q) = x_j^*(P)$  oder  $(x_j^*(P) = 1 \text{ und } y_j^*(Q) = 0)$ . Ansonsten erfolgt eine Änderung in Abhängigkeit von der Distanz  $d_j^*(P, Q)$  und der Differenz  $y_j^*(Q) - x_j^*(P)$ . Dabei fällt die Änderung umso stärker aus, je kleiner  $x_j^*(P)$ . Wenn  $x_j^*(P) < y_j^*(Q)$ , dann wird der Informationsgrad von  $x_j$  bezüglich P erhöht, ansonsten reduziert. Analoge Aussagen gelten für  $y_j^*(Q)$ .

Auch die anderen charakteristischen Werte von  $x_j$  und  $y_j$  werden beeinflusst. Sei  $T$  der aktuelle Zeitpunkt der Wechselwirkung. Für  $x_j$  ergeben sich folgende Änderungen (entsprechend für  $y_j$ ):

Fall 1:  $\emptyset H_2 HT(x_j, P)$  und  $HT(x_j, P)$  undefiniert, d. h.  $x_j$  ist nicht im Profil  $P$

Setze  $\emptyset H_2 HT(x_j, P) := \emptyset H_2 HT(y_j, Q)$  und  $HT(x_j, P) := T$ .

Fall 2:  $\emptyset H_2 HT(x_j, P)$  undefiniert,  $HT(x_j, P) < T$

Setze  $\emptyset H_2 HT(x_j, P) := T - HT(x_j, P)$  und  $HT(x_j, P) := T$ .

Fall 3:  $\emptyset H_2 HT(x_j, P)$  definiert,  $HT(x_j, P) < T$

Setze  $\emptyset H_2 HT(x_j, P) := (\emptyset H_2 HT(x_j, P) + T - HT(x_j, P)) / 2$  und  $HT(x_j, P) := T$ .

Falls  $HT(x_j, P) = T$  ändern sich die charakteristischen Werte nicht.

Im letzten Schritt definieren wir nun ein geeignetes Maß zur Bestimmung der Distanz zwischen Profilen.

## 4 Maß zur Bestimmung der Distanz von Profilen

Um die Ähnlichkeit von Profilen  $P$  und  $Q$  im Rahmen des Profil-Matching zu bestimmen, benötigen wir ein Distanzmaß<sup>25</sup>  $d(P, Q)$  innerhalb des Profil-Raumes. Hierzu definieren wir zunächst für jede (Attribut-)Dimension  $A$  ein separates aber allgemeines, d. h. von der Skalierung der Dimension unabhängiges, Distanzmaß  $d_A(P, Q)$ . Offensichtlich sollte  $d_A(P, Q)$  von sämtlichen Ausprägungen abhängen die  $P$  und  $Q$  für das Attribut  $A$  besitzen. Bei der Distanz-Berechnung ist zu berücksichtigen, dass Profile im Sinne des Profil-Matching als ähnlich gelten, wenn sie gemeinsame Attribut-Ausprägungen mit einem hohen Informationsgrad besitzen. Profile die nur Attribut-Ausprägungen mit einem Informationsgrad nahe bei 0 besitzen, sollten hingegen eine besonders hohe Distanz aufweisen<sup>26</sup>.

Für Attribut  $A$  seien wiederum Profil  $P$  die Ausprägungen  $x_1, \dots, x_s$  und Profil  $Q$  die Ausprägungen  $y_1, \dots, y_r$  mit positivem Informationsgrad zugeordnet. Alle Attribut-Ausprägungen in  $Q$  die nicht in  $P$  liegen, werden auch in  $P$  eingefügt, allerdings mit Informationsgrad 0. Entsprechend verfährt man mit Attribut-Ausprägungen in  $P$

<sup>25</sup> Der Begriff Maß wird hier nicht im streng mathematischen Sinne verwandt, sondern beschreibt allgemein eine Berechnungsvorschrift für einen Vergleichswert.

<sup>26</sup> Hier würde eine "Ähnlichkeit" ja nur im Fehlen von echten Attribut-Ausprägungen bestehen. Dies ist aber beim Profil-Matching gerade nicht erwünscht.

die nicht in  $Q$  liegen. Anschließend sortiert man die Ausprägungen so um, dass  $x_j = y_j$  für  $j = 1, \dots, n$ . Dabei bezeichne  $n$  erneut die Anzahl der gemeinsamen Attribut-Ausprägungen in Profil  $P$  und  $Q$ .

Basierend auf den gemäß Formel (12) definierten Distanzen zwischen  $P$  und  $Q$  bezüglich der Ausprägungen des Attributes  $A$ , ermitteln wir die informationsbewertende Attribut-Distanz wie folgt:

$$(13) \quad d_A(P, Q) := 1 - \frac{\sum_{j=1}^n (1 - d_{j*}(P, Q))}{\sum_{j=1}^n x_{j*} + \sum_{j=1}^n y_{j*} - \sum_{j=1}^n (1 - d_{j*}(P, Q))}$$

Wie man sich leicht überlegt gilt  $d_A(P, Q) \in [0, 1]$ , mit  $d_A(P, Q) = 0$  gdw.  $P$  und  $Q$  dieselben Ausprägungen für Attribut  $A$  besitzen (d. h.  $s = r = n$ ) und allen diesen Ausprägungen der Informationsgrad 1 zugeordnet ist.

Die Attribut-Distanz wird also in Anlehnung an das – aus dem Information Retrieval bekannte – Ähnlichkeits-Maß von Tanimoto [Kohonen 97] definiert. Anders als viele andere entsprechend adaptierte Distanz- oder Ähnlichkeitsmaße vergrößert dieses Maß wunschgemäß die Ähnlichkeit zwischen Profilen die gemeinsame Ausprägungen mit einem hohen Informationsgrad besitzen und senkt die Ähnlichkeit zwischen Profilen, die nur Ausprägungen mit einem niedrigen Informationsgrad besitzen.

Der Gesamtabstand  $d(P, Q)$  zwischen  $P$  und  $Q$  wird schließlich als mittlere Attribut-Distanz zwischen  $P$  und  $Q$  festgelegt:

$$(14) \quad d(P, Q) := \frac{1}{k} \sum_{i=1}^k d_{A_i}(P, Q) .$$

Dabei kann es u. U. sinnvoll sein nur diejenigen Attribut-Distanzen zu berücksichtigen, die unterhalb eines vorzugebenden Schwellwertes liegen<sup>27</sup>.

## 5 Ausblick

Es wurde ein Konzept für ein Profilingsystem vorgestellt, in dessen Zentrum ein Profilmodell steht, bei dem jeder Attribut-Ausprägung ein Informationsgrad zuge-

---

<sup>27</sup> In diesem Fall sollte allerdings gewährleistet sein, dass beim Profil-Influencing ebenfalls nur diejenigen Attribute angepasst werden, die eine geringe Attribut-Distanz besitzen.

ordnet ist, der sowohl beim Profil-Influencing, als auch beim Profil-Matching geeignet berücksichtigt wird. Dadurch ist es möglich bei der Profilbildung auch mit Informationen geringer Validität bzw. mit Prognosen zu arbeiten. Zusammen mit dem Ansatz, neben Nutzer-Profilen auch Item-Profile zu verwenden und diese sich wechselseitig beeinflussen zu lassen, liefert dies eine Lösung für das Kaltstartproblem von Systemen zur automatisierten Personalisierung mittels Profilen.

Noch offen ist die Frage, ob sich das Konzept auch auf Systeme übertragen lässt, die mit hierarchischen Relationen zwischen Attribut-Ausprägungen arbeiten. Für den praktischen Einsatz bleibt ferner zu klären, wie die gleichzeitige Wechselwirkung zwischen mehr als zwei Profilen, sowie die Verschmelzung von Profilen zu handhaben sind. In der Vergangenheit wurden Teile des Konzeptes in einem Recommender-System des ehemaligen Internet-Portals „yoolia“ implementiert. Eine vollständige Realisierung und Evaluierung auf vergleichbarer Basis [Herlocker et al. 00] ist für die Zukunft im Rahmen von Diplom- bzw. Masterarbeiten sowie einem Forschungsprojekt geplant, für das noch Partner aus der Wirtschaft gesucht werden.

## 6 Literatur

- [Breese et al. 98] Breese, J. S., Heckermann, D., Kadie, C. „Empirical Analysis of Predictive Algorithms for Collaborative Filtering“, Proc. of Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, Morgan Kaufmann, 1998
- [Dunn et al. 97] Dunn, M., Gwertzman, J., Layman, A., Partovi, H. „Privacy and Profiling on the Web“, W3C Working Draft, 1997
- [Galliat 99] Galliat, T. „Clustering Data of Different Information Levels“, International Data Analysis Conf. (IDA), Innsbruck, 2000 (available as ZIB-Preprint SC-99-42, Berlin, 1999)
- [Herlocker et al. 00] Herlocker, J., Konstan, J., Terveen, L., Riedl, J. „Evaluating Collaborative Filtering Recommender Systems“, ACM Transactions on Information Systems 22, p. 5-53, ACM Press, 2004
- [Kobsa 04] Kobsa, A. „Adaptive Verfahren – Benutzermodellierung“, In: Kuhlen, R., Seeger, T., Strauch, D. (Hrsg.) „Grundlagen der praktischen Information und Dokumentation“, 5. Auflage, K.G. Saur, München, 2004
- [Kohonen 97] Kohonen, T. „Self-Organizing Maps“, 2nd Ed., p. 16-17, Springer, Berlin, 1997
- [Massa et al. 04] Massa, P., Avesani, P. „Trust-aware Collaborative Filtering for Recommender Systems“, Proc. of the Intern. Conf. on Cooperative Information Systems (CoopIS), Larnaca, 2004
- [Schein et al. 02] Schein A. I., Popescul A., Ungar L. H., Pennock D. M., „Methods and Metrics for Cold-Start Recommendations“, Proc. of the 25<sup>th</sup> Annual Intern. ACM SIGIR

- Conf. on Research & Development on Information Retrieval (SIGIR02), Tampere, Finland, 2002
- [Terveen et al. 01] Terveen, L., Hill, W. „Beyond Recommender Systems: Helping People Help Each Other“, In Carroll, J. (ed.), HCI in the New Millennium. Addison Wesley, 2001.
- [Ungar et al. 98] Ungar, L. H., Foster, D. P. „Clustering Methods für Collaborative Filtering“, Proc. of the Workshop on Recommendation Systems. AAAI Press, Menlo Park California, 1998
- [van Meteren et al. 00] van Meteren, R., van Someren, M. „Using Content-Based Filtering for Recommendation“, MLnet / ECML2000 Workshop, Barcelona, 2000
- [Wang et al. 06] Wang, J., de Vries, A. P., Reinders, M. J. T. „Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion“, Proc. of the 29<sup>th</sup> Annual Intern. ACM SIGIR Conf. on Research & Development on Information Retrieval (SIGIR06), Seattle, 2006



# Die Tagungsbände des Internationalen Symposiums für Informationswissenschaft in szientometrischer Analyse\*

*Maria Baumgartner, Christian Schlögl*

Universität Graz

Institut für Informationswissenschaft und Wirtschaftsinformatik

Universitätsstrasse 15/F3, 8010 Graz

*marie79b@hotmail.com, christian.schloegl@uni-graz.at*

## **Zusammenfassung**

Anlässlich der zehnten Durchführung eines Internationalen Symposiums für Informationswissenschaft (ISI) sollen die Proceedings der bisherigen Tagungen einer szientometrischen Analyse unterzogen und auf dieser Basis eine Bestandsaufnahme des Symposiums vorgenommen werden. Die Analysen erfolgen auf Ebene der Artikel, Autoren, Institutionen und Referenzen. Anhand ausgewählter Indikatoren wird die Tagungsreihe in Bezug auf Internationalität, wissenschaftliche Orientierung und „identitätsstiftende“ Wirkung für die deutschsprachige Informationswissenschaft kritisch hinterfragt.

## **Abstract**

The International Symposium on Information Science takes place the tenth time this year. On this anniversary the proceedings of the previous symposia are investigated by means of a scientometric study. The analyses are conducted at the level of articles, authors, institutions and references. On the basis of selected indicators the conference series is examined with regards to internationality, research focus and its identity-creating effects on German-speaking information science.

## **I Einleitung**

Das Internationale Symposium für Informationswissenschaft (ISI) ist die zentrale informationswissenschaftliche Tagung in den deutschsprachigen Ländern. Sie wird

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 43-59.

vom Hochschulverband Informationswissenschaft (HI) in Kooperation mit der informationswissenschaftlichen Hochschuleinrichtung, welche die jeweilige Tagung organisiert, veranstaltet. Auf der Homepage des Hochschulverbands Informationswissenschaft wird das Selbstverständnis der Tagungsreihe wie folgt definiert: „Die Symposien stehen stellvertretend für die Entwicklung der Informationsgesellschaft. Sie beschäftigen sich stets sowohl mit den zentralen Aufgaben als auch wechselnden Anforderungen, die an die Informationswissenschaft gestellt werden und den Schwierigkeiten, die sich daraus ergeben.“ [HI 06]

<i>Jahr</i>	<i>Tagungsort</i>	<i>Herausgeber</i>	<i>Titel des Tagungsbandes</i>
1990	Konstanz	Josef Herget Rainer Kuhlen	Pragmatische Aspekte beim Entwurf und Betrieb von Informationssystemen
1991	Oberhof Thüringen	Harald Killenberg Rainer Kuhlen Hans-Jürgen Manecke	Wissensbasierte Informationssysteme und Informationsmanagement
1992	Saarbrücken	Harald H. Zimmermann	Mensch und Maschine – informationelle Schnittstellen der Kommunikation
1994	Graz	Wolf Rauch Franz Strohmeier Harald Hiller Christian Schlögl	Mehrwert von Information – Professionalisierung der Informationsarbeit
1996	Berlin	Jürgen Krause Matthias Herfurth Jutta Marx	Herausforderungen an die Informationswissenschaft: Informationsverdichtung, Informationsbewertung und Datenvisualisierung
1998	Prag	Harald H. Zimmermann Volker Schramm	Knowledge Management und Kommunikationssysteme: Workflow Management, Multimedia, Knowledge Transfer
2000	Darmstadt	Gerhard Knorz Rainer Kuhlen	Informationskompetenz – Basiskompetenz in der Informationsgesellschaft
2002	Regensburg	Rainer Hammwöhner Christian Wolff Christa Womser-Hacker	Information und Mobilität – Optimierung und Vermeidung von Mobilität durch Information
2004	Chur	Bernard Bekavac Josef Herget Marc Rittberger	Information zwischen Kultur und Marktwirtschaft

*Tabelle 1: Bisher stattgefundenene Tagungen mit Titel und Herausgeber des jew. Tagungsbandes*

Tabelle 1 gibt einen Überblick über Tagungsort, Herausgeber und Generalthema der bisherigen Symposien. Wie zu sehen ist, fand ISI nicht nur an Standorten informationswissenschaftlicher Hochschuleinrichtungen in Deutschland, sondern auch



in Graz (ISI 1994), Prag (ISI 1998) und in Chur (ISI 2004) statt. In Köln wird das Symposium zum zehnten Mal ausgerichtet. Anlässlich dieses Jubiläums sollen die Proceedings der bisherigen Tagungen einer szientometrischen Analyse unterzogen und auf dieser Basis eine Bestandsaufnahme des Symposiums vorgenommen werden.

## **2 Zeitschriftenszientometrie**

Van Raan definiert Szientometrie in einem State-of-the-Art-Artikel folgendermaßen: “Scientometric research is devoted to quantitative studies of science and technology” ([Van Raan 97 : 205]). Die Szientometrie befasst sich also mit der Anwendung exakter Messmethoden bei der Untersuchung der Wissenschaft.

Nachdem Zeitschriften und Tagungen in den meisten wissenschaftlichen Disziplinen als zentrales Instrument zur Verbreitung der wissenschaftlichen Erkenntnisse dienen, überrascht es nicht, wenn Periodika selbst Gegenstand vieler szientometrischer Studien sind. Die Bedeutung der Zeitschriftenszientometrie wird nicht zuletzt dadurch dokumentiert, dass es mit den Journal Citation Reports (JCR) eine eigene Datenbank gibt, in der die wichtigsten Kennzahlen der im Web of Science (WOS) erfassten Periodika abgerufen werden können (vgl. [White & McCain 89 : 149]).

In der Literatur gibt es mittlerweile einen reichen Fundus an Zeitschriftenanalysen. Naturgemäß wurden die meisten Studien in den Informationswissenschaften durchgeführt (vgl. z. B. [Schubert 01] oder [Koehler et al. 01]). Deutschsprachige Periodika waren seltener Objekt derartiger Analysen. Einer der Hauptgründe dürfte neben der geringeren internationalen Orientierung darin liegen, dass diese Zeitschriften oft nicht in bibliographischen Datenbanken erfasst sind und die Datenerhebung/-analyse daher meist sehr aufwändig ist. Die in den letzten Jahren umfassendste Studie dieser Art untersuchte 40 internationale und die zehn bedeutendsten deutschsprachigen Zeitschriften des Bibliotheks-, Informations- und Dokumentationswesens sowie der Informationswissenschaft (siehe [Grazia Colonia 02a], [Grazia Colonia 02b], [Schloegl & Stock 04], [Schloegl & Stock 06]). In Detailstudien wurden einzelne deutschsprachige Zeitschriften einer vertiefenden Analyse unterzogen (vgl. [Dresel 03], [Juchem 02], [Juchem 04], [Juchem, Schlögl & Stock 06]). Ähnlich gelagerte Untersuchungen liegen auch über Zeitschriften anderer Disziplinen vor (zum Beispiel [Albrecht 99] oder [Resch & Schlögl 04], um nur einige wenige zu nennen).

### 3 Vorgehensweise und Methode

Der Web of Science oder die auf Zeitschriftenebene aggregierten Daten der Journal Citation Reports (JCR) (vgl. [Stock 01]) gelten als die klassischen Instrumente bei der Durchführung von Zeitschriftenanalysen. Erst in den letzten Jahren wurde mit Scopus ein ernstzunehmendes Konkurrenzprodukt auf den Informationsmarkt gebracht. Da die ISI-Proceedings in diesen beiden Datenbanken nicht enthalten sind und zudem eine Referenzanalyse durchgeführt werden sollte, mussten die Daten manuell erhoben werden. Um den Aufwand für die Auswertungen so gering wie möglich zu halten, wurden die einzelnen Artikel in eine eigens für diesen Zweck erstellte Access-Datenbank eingegeben.

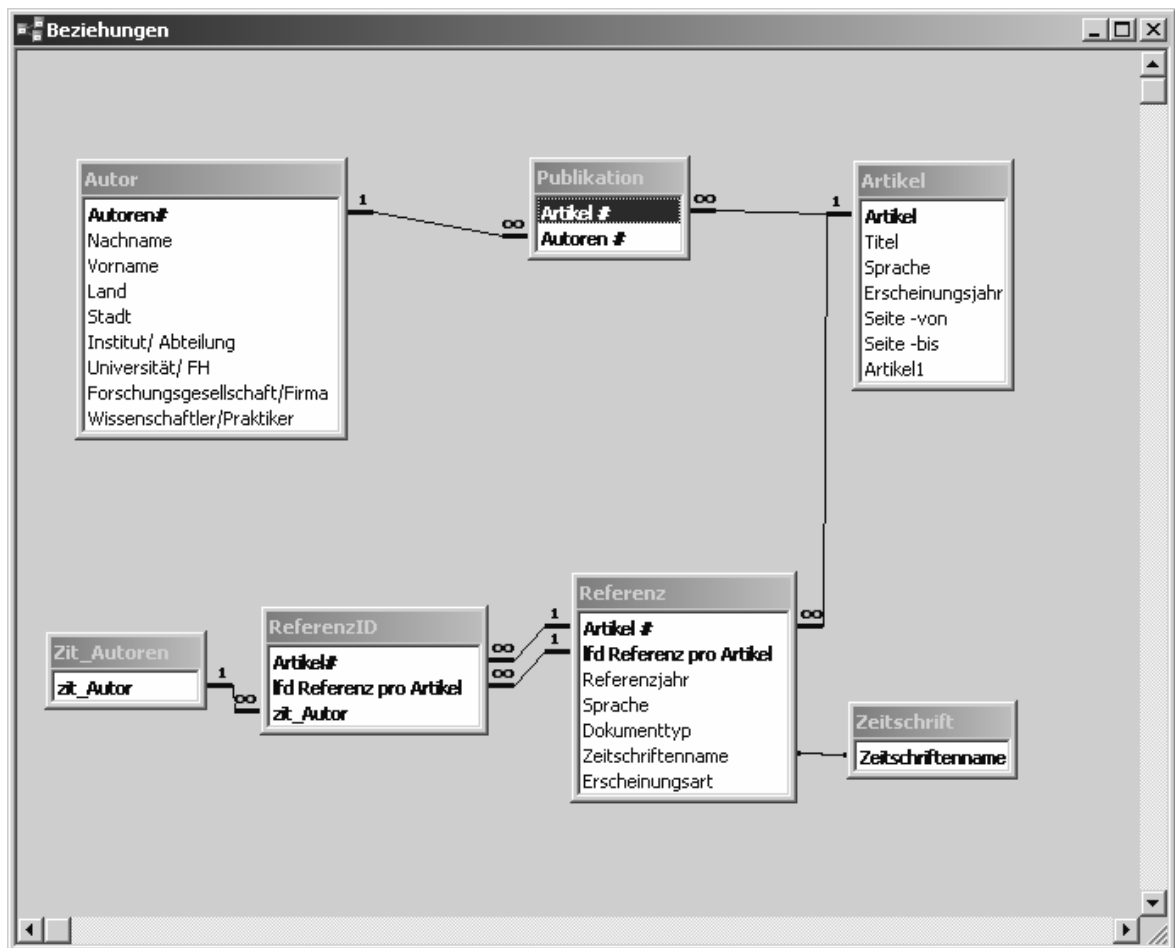


Abbildung 1: Struktur der Access-Datenbank [Baumgartner 05]

Abbildung 1 zeigt den Aufbau der Datenbank. Wie zu sehen ist, sind aufgrund der Datenbankstruktur auch Auswertungen möglich, die in herkömmlichen bibliographischen Datenbanken nicht erstellt werden können. Beispielsweise werden im Gegensatz zum Web of Science alle Autoren einer zitierten Quelle erfasst, wodurch nicht nur Erstautoren in Zitations-Rankings berücksichtigt werden. Weiter wird

von einem Autor auch festgehalten, ob er an einer Universität oder an einer Fachhochschule beschäftigt ist. Dadurch kann untersucht werden, in welchem Maße Fachhochschulen in die ISI-Tagungen eingebunden sind.

Für die Datenerfassung wurden komfortable Eingabeformulare verwendet. Die Auswertungen erfolgten unter Verwendung von Berichten, die auf zuvor erstellten Abfragen basierten. (Die Details der Datenbankanwendung können ebenso wie die detaillierten Ergebnisse der einzelnen Auswertungen in [Baumgartner 05] nachgelesen werden.) Bei komplexeren Auswertungen, die mit den Access-Assistenten nicht gelöst werden konnten, wurden die gewünschten Daten exportiert und mit MS Excel weiterverarbeitet.

Die Analysen erfolgten auf folgenden Ebenen:

- Artikel
- Autoren
- Institutionen und
- Referenzen.

## **4 Ergebnisse**

In diesem Kapitel werden die Ergebnisse, den Analyseebenen entsprechend, vorgestellt.

### **4.1 Artikelanalyse**

Bei der Artikelanalyse werden folgende Sachverhalte untersucht:

- Anzahl der Artikel je Tagungsband
- durchschnittliche Anzahl der Autoren je Artikel
- Publikationssprache
- durchschnittliche Anzahl der Referenzen je Artikel.

Die neun Tagungsbände umfassten insgesamt 352 Artikel, was einem Durchschnitt von 39 Artikeln je Tagungsband entspricht. Wie aus Tabelle 2 zu entnehmen ist, weichen die einzelnen Proceedings relativ stark von diesem Mittelwert ab. Am meisten Beiträge (59) enthielt der Proceedings der Tagung in Graz, am wenigsten (18) jener der Darmstädter Tagung. Bei Letzterer wurde allerdings ein weiterer Tagungsband (siehe [Knorz & Otto 00]) herausgegeben, der ausschließlich die Beiträge des anwendungsorientierten Programms enthielt. Im Gegensatz dazu schwankt das Ausmaß der Koautorenschaft zwischen den einzelnen Proceedings wesentlich weniger. Im Durchschnitt (über alle neun Tagungen) wird ein Aufsatz

von 1,6 Autoren verfasst. Insgesamt wirkten in nur 25 Artikeln (das sind 7,1% aller Artikel) mehr als zwei Autoren mit. Das relativ geringe Ausmaß an gemeinsam verfassten Publikationen ist in der Informationswissenschaft, wie auch in vielen anderen sozialwissenschaftlichen Disziplinen, durchaus üblich.

Jahr	Tagungsort	Artikel				Autoren		Referenzen	
		ges.	dt.	engl.	engl. %	ges.	je Artikel	ges.	je Artikel
1990	Konstanz	46	40	6	13%	58	1,3	829	18,0
1991	Oberhof	43	40	3	7%	61	1,4	493	11,5
1992	Saarbrücken	34	28	6	18%	50	1,5	572	16,8
1994	Graz	59	43	16	27%	88	1,5	1057	17,9
1996	Berlin	35	35	25	29%	60	1,7	522	14,9
1998	Prag	45	26	19	42%	68	1,5	611	13,6
2000	Darmstadt	18	13	5	28%	32	1,8	318	17,7
2002	Regensburg	34	28	6	18%	82	2,4	469	13,8
2004	Chur	38	35	3	8%	56	1,5	608	16,0
Summe		352	278	74	21%	555	1,6	5479	15,6

Tabelle 2: Kennzahlen der Artikelanalyse (je Tagungsband)

Auch die Länge der Literaturliste zeigt zwischen den einzelnen Tagungen keine allzu großen Unterschiede. Durchschnittlich hat ein Artikel 15,6 Referenzen. Im Vergleich zu den führenden internationalen informationswissenschaftlichen Zeitschriften (zum Beispiel JASIST: 43; Journal of Documentation: 55; Information Processing & Management: 34 Referenzen je Artikel) ist dieser Wert relativ gering (die Werte wurden aus den JCR des Jahres 2005 ermittelt). Bei der Gegenüberstellung mit den Proceedings of the ASIST Annual Meeting (ca. 20 Referenzen pro Artikel – in den JCR konnten die entsprechenden Werte aufgrund von fehlenden Daten nur für die Jahre 1997, 1998 und 2002 ermittelt werden) ergibt sich ein deutlich geringerer Unterschied.

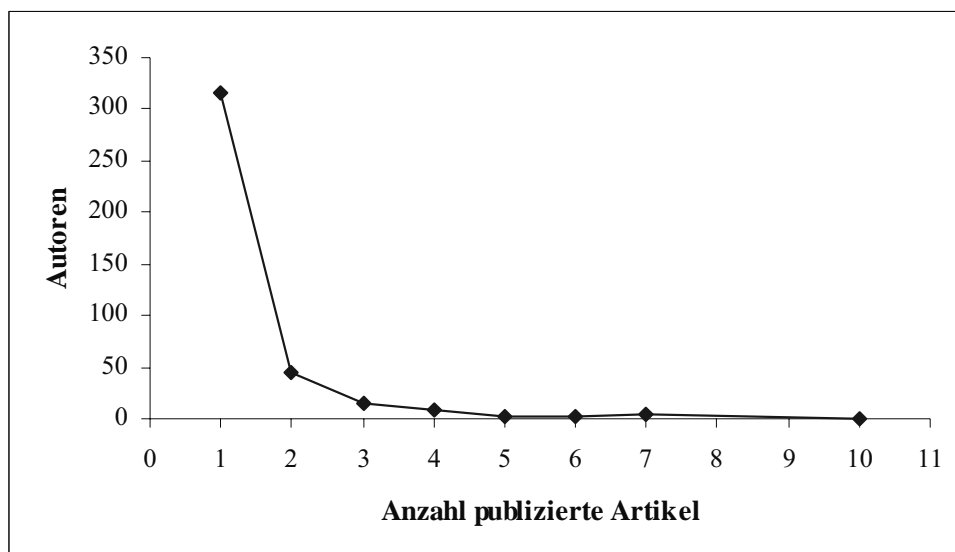
Im Schnitt wurde jeder fünfte Artikel in englischer Sprache verfasst, wobei es hier wieder größere Unterschiede zwischen den einzelnen Tagungen gibt. Naturgemäß am meisten englische Beiträge enthielt der Proceedings der Prager Tagung. Beim zweiten Symposium in Oberhof lagen hingegen nur 7% der veröffentlichten Aufsätze in Englisch vor.

#### 4.2 Autoren- und Institutionenanalyse

Im Rahmen der Autoren- und Institutionenanalyse wurden folgende Auswertungen durchgeführt:

- Anzahl der verfassten Artikel (Häufigkeitsverteilung und Ranking)
- Herkunft der Autoren
- Institutionszugehörigkeit der Autoren
- Verhältnis von Praktikern und Wissenschaftlern.

In den neun Tagungsbänden haben 396 (verschiedene) Autoren publiziert. Wie aus Abbildung 2 hervorgeht, ist die Autorenproduktivität extrem schief verteilt. Demnach haben zirka 80% der Autoren nur einen Aufsatz veröffentlicht. Lediglich 4,8% aller Autoren (19 Personen) haben mehr als drei Artikel verfasst.



*Abbildung 2: Autorenproduktivität*

In Tabelle 3 sind diese 19 Autoren nach der Anzahl der in den ISI-Tagungsbänden veröffentlichten Artikel gereiht. Wie zu sehen ist, wurden die meisten Artikel (10 Aufsätze) von Christian Wolff verfasst. Die nächst gereihten Autoren haben bereits deutlich weniger Beiträge (7 Aufsätze) vorzuweisen. Werden bei der Artikelzählung Mehrfachautorenschaften aliquot berücksichtigt (zum Beispiel bei vier Autoren mit einem Viertel), so ergibt sich ein zum Teil abweichendes Ranking. Bei diesem bringen es Schlögl, Wolff und Yetim auf fast gleich viele Beiträge. Interessant ist schließlich noch die Frage, in wie vielen Proceedings ein Autor mit mindestens einem Aufsatz beigetragen hat, also nach der „Stammkundschaft“ aus Autorensicht. Hier wird das Ranking wiederum von Christian Wolff angeführt, der an acht Tagungsbänden mit mindestens einem Beitrag als (Ko-)Autor mitgewirkt hat. Nur acht Autoren waren in mehr als der Hälfte der bisherigen Proceedings mit (mindestens) einem Artikel vertreten.

Wie ebenfalls aus Tabelle 3 zu entnehmen ist, kommen die Autoren mit den meisten Veröffentlichungen ausschließlich aus dem deutschsprachigen Raum. Auch bei

der Berücksichtigung aller Autoren zeigt sich die Dominanz der deutschsprachigen Länder. Aus diesen kommen durchschnittlich 81% der Autoren, der Großteil davon (insgesamt 65%) aus Deutschland (siehe Abbildung 3).

Rang	Nachname	Vorname	Aufsätze	Aufsätze mit Berücks. v. Mf.-Autorensch.		Aufsätze in Proceedings	
					Rang		Rang
1.	Wolff	C.	10	5	37/60	2.	8
2.	Kuhlen	R.	7	5	1/6	4.	7
2.	Mandl	T.	7	3	11/12	7.	3
2.	Schlögl	C.	7	5	13/20	1.	6
2.	Womser-Hacker	C.	7	3	1/6	11.	5
6.	Englmeier	K.	6	5		5.	6
6.	Herget	J.	6	4	5/6	6.	6
6.	Yetim	F.	6	5	1/2	3.	6
9.	Rittberger	M.	5	3	1/3	10.	5
9.	Semar	W.	5	3	2/3	8.	4
11.	Bekavac	B.	4	1	1/2	19.	3
11.	Griesbaum	J.	4	2		16.	2
11.	Hammwöhner	R.	4	3	1/2	9.	4
11.	Lehner	F.	4	3		12.	4
11.	Milchrahm	E.	4	3		12.	4
11.	Panyr	J.	4	2	5/6	14.	4
11.	Teufel	S.	4	1	2/3	18.	2
11.	Thiel	U.	4	2		16.	4
11.	Weitzendorf	T.	4	2	1/2	15.	4

Tabelle 3: Publierte Artikel (ohne und mit Berücksichtigung von Mehrfachautorenschaften sowie Anzahl der Tagungsbände, an denen jew. Autor mit mindestens einem Aufsatz mitgewirkt hat)

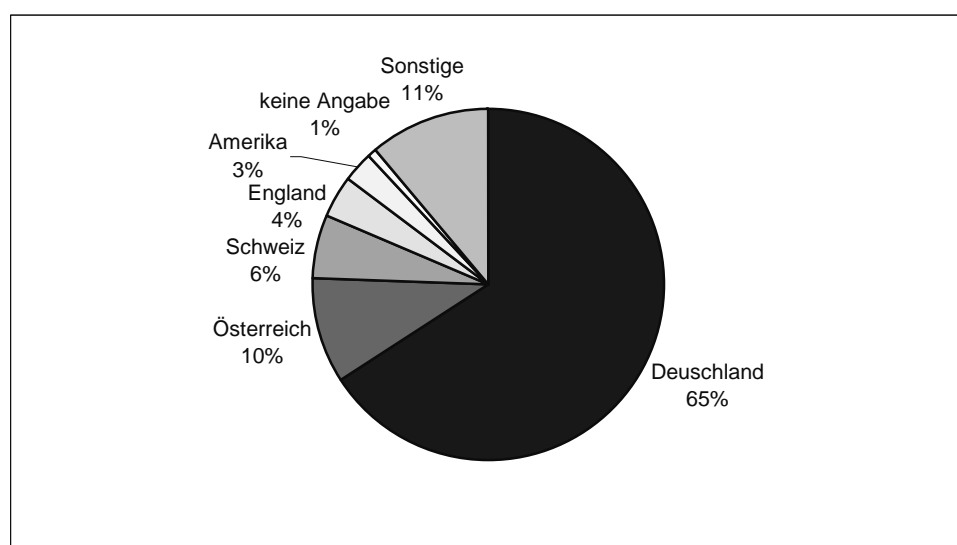


Abbildung 3: Länder der Autoren

Wenig überraschend ist, dass die Universität Konstanz bei den ersten neun Tagungen mit Abstand am meisten Autoren stellte (66), gefolgt von den Universitäten Graz (39 Autoren), Regensburg (26 Autoren) und Hildesheim (19 Autoren), dem IZ Sozialwissenschaften (19 Autoren) und der Universität des Saarlandes (17) (siehe Tabelle 4). Etwas unerwartet war hingegen, dass sich im Ranking der am meist genannten Institutionen keine einzige Fachhochschule befindet. Auch wenn man die Institutszugehörigkeit aller Autoren berücksichtigt, ergibt sich kein wesentlich differenzierteres Bild. Demnach beträgt der Anteil der Fachhochschulen nur 6%. Das Gros der Autoren stellen Universitäten mit 57%. Sonstige Institutionen haben mit 36% einen relativ hohen Anteil, wovon Forschungsgesellschaften zirka zwei Drittel sowie Unternehmen und andere Praxiseinrichtungen ungefähr ein Drittel stellen.

<i>Rang</i>	<i>Institution</i>	<i>Anzahl</i>	<i>Typ</i>
1	Uni Konstanz	66	Universität
2	Uni Graz	39	Universität
3	Uni Regensburg	26	Universität
4	Uni Hildesheim	19	Universität
4	IZ Sozialwissenschaften	19	Sonstige
6	Uni des Saarlandes	17	Universität
7	Siemens AG	14	Sonstige
8	Gesellschaft für Mathematik und Datenverarbeitung (GMD)	13	Sonstige
9	Uni Leipzig	12	Universität
9	Fraunhofer Institut	12	Sonstige
11	Uni Zürich	10	Universität

*Tabelle 4: Institutionszugehörigkeit der Autoren (mindestens zehn Nennungen)*

Was die berufliche Orientierung der Autoren betrifft, handelt es sich eindeutig um eine wissenschaftliche Tagung. Im Durchschnitt waren 74% aller Autoren Wissenschaftler (bzw. Lehrende an einer Fachhochschule). Den größten Praktikeranteil hatte das Symposium in Chur mit 34%.

### **4.3 Referenzanalyse**

Im Zuge der Referenzanalyse werden folgende Aspekte behandelt:

- Anzahl der erhaltenen Zitate je Autor (Ranking)
- Sprache der Referenzen
- Halbwertszeit
- Dokumenttyp
- Erscheinungsform (gedruckt/digital/beides)

- zitierte Zeitschriften (Ranking) und Selbstreferenzrate des ISI-Tagungsbandes.

Wie bereits erwähnt, wurden insgesamt 5479 Referenzen erfasst. Diese entfielen auf 5202 (verschiedene) Autoren. Mit Abstand am öftesten zitiert (127-Mal) wurde Rainer Kuhlen, nicht einmal halb so viele Zitate (55) erhielt der zweitgeriehte Autor (Jürgen Krause) mit 55 Zitaten (siehe Tabelle 5). Im Gegensatz zum Autoren-Top-Ranking sind unter den am meisten zitierten Autoren auch einige aus dem nichtdeutschsprachigen Ausland stammende und primär im Bereich des Information Retrieval ausgewiesene Wissenschaftler (Belkin, Conklin, Pejtersen, Salton) zu finden.

Rang	Autor	Zitate
1	Kuhlen R.	127
2	Krause J.	55
3	Womser-Hacker C.	38
4	Mandl T.	36
5	Yetim F.	34
6	Fuhr N.	33
7	Wolff C.	30
8	Hammwöhner R.	29
9	Belkin N. J.	27
10	Conklin J.	22
11	Herget J.	21
12	Pejtersen A. M.	20
12	Salton G.	20
12	Thiel U.	20

Tabelle 5: Autoren die am öftesten zitiert wurden (mindestens 20 Zitate)

Jahr	deutschspr. Quellen		englischspr. Quellen		Sonstige	Gesamt	Halbwertszeit
	abs.	rel.	abs.	rel.			
1990	361	44%	460	55%	8	829	4,8
1991	312	63%	177	36%	4	493	5,3
1992	205	36%	367	64%	0	572	3,5
1994	414	39%	636	60%	7	1057	5,6
1996	206	39%	311	60%	5	522	5,0
1998	221	36%	355	58%	35	611	2,9
2000	141	44%	177	56%	0	318	2,8
2002	192	41%	277	59%	0	469	3,4
2004	257	42%	351	58%	0	608	2,8
	2309	42%	3111	57%	59	5479	3,7

Tabelle 6: Sprache und Halbwertszeit der Referenzen



Da auch die Sprache der zitierten Quellen festgehalten wurde, ließ sich einfach ermitteln, ob die Autoren der ISI-Tagungen internationale Forschungsergebnisse in einem stärkeren Maße aufgreifen. Dies dürfte auch tatsächlich der Fall sein. Immerhin 57% aller Referenzen verweisen auf englischsprachige Quellen. Nur im Proceedings des zweiten Symposiums in Oberhof wurden großteils deutschsprachige Quellen zitiert (siehe Tabelle 6).

Die Halbwertszeit der Referenzen gilt als ein Indikator für das Alter der verwendeten Quellen. Sie ist bestimmt durch jenen Zeitraum, in dem die Hälfte (Median) der zitierten Literatur, absteigend nach dem Erscheinungsjahr sortiert, hineinfällt. Für die neun Tagungsbände wurde ein Mittelwert von 3,7 Jahren berechnet. Die führenden informationswissenschaftlichen Periodika haben hingegen eine deutlich langfristige Perspektive, was die in ihnen verwendeten Quellen betrifft (JASIST: 7,5 Jahre; Journal of Documentation: 9,6 Jahre; Information Processing & Management: 6,9 Jahre (JCR 2005)). Da die Halbwertszeit auch von der Textsorte abhängig ist – so sollten Tagungsbände in der Regel aktuellere Beiträge enthalten und folglich eine niedrigere Halbwertszeit haben –, wurde auch ein Vergleich mit den Proceedings of the ASIST Annual Meeting angestellt. Für diesen war im Jahr 2005 ebenfalls ein sehr hoher Wert von 7,8 (JCR 2005) ausgewiesen. Die Werte für die Jahre 2001 (5,4 Jahre) und 2002 (5,5 Jahre) (jene Jahre, in denen die entsprechenden Daten in den JCR verfügbar waren) dürften aber eher dem langjährigen Durchschnitt entsprechen.

Interessant ist die Entwicklung der Halbwertszeit im Zeitablauf (siehe Tabelle 6). So lag die Halbwertszeit in den letzten vier Tagungsbänden deutlich unter den Werten, die für die ersten fünf Proceedings berechnet wurden. Eine mögliche Ursache könnte auch darin liegen, dass sich die Zitiergewohnheiten in diesem Zeitraum geändert haben. Diese sollen anhand der Zusammensetzung der zitierten Dokumenttypen und deren Erscheinungsform (print / digital / beides) als nächstes untersucht werden.

Wie in Abbildung 4 deutlich zu erkennen ist, wurden ab der Tagung im Jahr 2000 deutlich weniger Monographien zitiert. Im Gegensatz dazu gewannen Web-Seiten seit 1996 an Bedeutung. Bei der letzten Tagung (2004) machten sie sogar einen Anteil von 28 % aus, der nicht nur zu Lasten der Monographien sondern auch zu Lasten von Aufsätzen in Herausgeberwerken und Grauer Literatur ging. Aufgrund der relativ geringen Anzahl der Beiträge (und somit auch der Referenzen) je Proceedings könnten die Ergebnisse aber mitunter verzerrt sein.

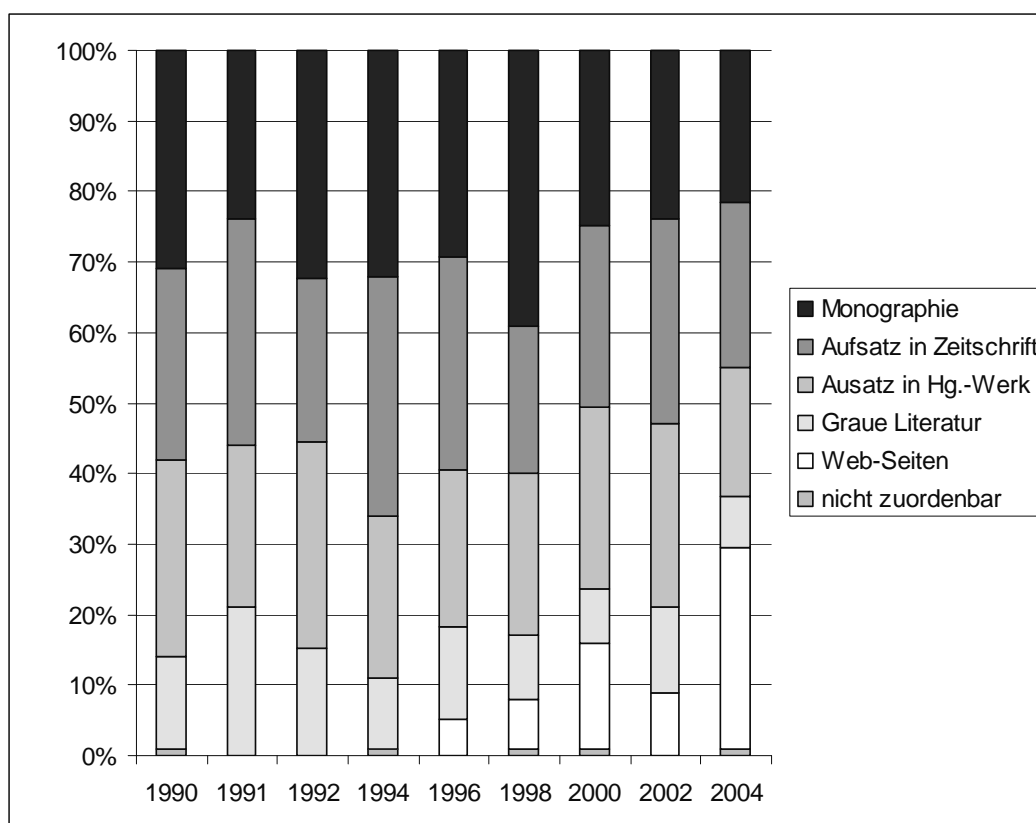


Abbildung 4: Anteile der Dokumenttypen an den Referenzen

Nicht zuletzt aufgrund des Internet hat sich die Erscheinungsform der zitierten Quellen in den 14 Jahren, in denen die neun Tagungsbände erschienen sind, sukzessive geändert (siehe Tabelle 7). Wurde in den ersten drei Tagungsbänden fast ausschließlich auf gedruckte Quellen verwiesen, so wurden ab Mitte der 90-er Jahre in einem zunehmenden Maße digitale Quellen zitiert. Diese machten beim letzten Symposium sogar 30% aus. Dieser Wert dürfte aber auch aufgrund der Spezifika einzelner Tagungsbeiträge etwas überhöht sein.

	1990	1991	1992	1994	1996	1998	2000	2002	2004	Mittel
Print	97,8%	98,0%	99,7%	95,2%	90,0%	89,7%	75,8%	84,9%	65,8%	89,9%
Digital	0,2%	0,2%	0,2%	1,3%	5,9%	3,1%	20,1%	12,6%	30,1%	7,4%
beides	1,9%	1,8%	0,2%	3,5%	4,0%	2,3%	4,1%	2,6%	4,1%	2,7%

Tabelle 7: Anteile der Erscheinungsformen an den Referenzen

Mit insgesamt 1528 Verweisen stellen Zeitschriftenaufsätze mit durchschnittlich 28% den am öftesten zitierten Dokumenttyp dar. Diese sollen nun einer weiteren Analyse unterzogen werden.

Diese Verweise entfallen auf 525 verschiedene Zeitschriften, was ein gewisses Indiz für die Interdisziplinarität der Informationswissenschaft im Allgemeinen und der Symposien im Besonderen darstellt. Tabelle 8 zeigt die am öftesten zitierten Zeit-

schriften. Wie zu sehen ist, handelt es sich dabei größtenteils um Zeitschriften aus dem Bereich von Informationswissenschaft und Informatik, im Top-Ranking aber ebenso enthalten sind Betriebswirtschaftslehre- und Wirtschaftsinformatik- bzw. MIS-Zeitschriften. Das Zeitschriften-Ranking wird von den beiden renommierten informationswissenschaftlichen Zeitschriften „JASIST“ (62 Zitate) und „Information Processing & Management“ (53 Zitate) angeführt. Für die deutschsprachige Informationswissenschaft eine gewisse Bedeutung hat die NfD (neuerdings IWP) (ebenfalls 53 Zitate), dies nicht zuletzt auch deshalb, weil es keine deutschsprachige informationswissenschaftliche Zeitschrift im eigentlichen Sinn gibt. 48 Referenzen verweisen auf Beiträge in (anderen) ISI-Tagungsbänden. Setzt man diese Zahl mit allen Referenzen (5479) in Beziehung, so ergibt sich eine Zeitschriften-Selbstreferenzrate von 0,9%. Nicht einmal jede hundertste Referenz bezieht sich auf eine andere Publikation in den ISI-Tagungsbänden.

Rang	Zeitschriftenname	Disziplin	Zitate	Sprache
1	JASIST	Infowiss.	62	englisch
2	Information Processing & Management	Infowiss.	53	englisch
2	NfD (IWP)	Infowiss.	53	deutsch
4	CACM – Communications of the ACM	Informatik	49	englisch
5	IM – Information Management & Consulting	MIS/WIN	48	deutsch
5	Tagungsband des ISI	Infowiss.	48	deutsch
7	Informatik Spektrum	Informatik	26	deutsch
8	Journal of Documentation	Infowiss.	22	englisch
9	Harvard Business Review	BWL	21	englisch
10	Journal of Information Science	Infowiss.	19	englisch
11	International Journal of Man-Machine Studies	Informatik	18	englisch
12	ACM Transactions Office Information Systems	Informatik	17	englisch
12	Wirtschaftsinformatik	MIS/WIN	17	deutsch
14	Scientometrics	Infowiss.	16	englisch
14	Sloan Management Review	BWL	16	englisch
16	ARIST	Infowiss.	15	englisch
16	IEEE Computer	Informatik	15	englisch
18	MIS Quarterly	MIS/WIN	14	englisch
19	Byte	Informatik	13	englisch
20	Artificial Intelligence	Informatik	12	englisch
20	Informatik – Forschung und Entwicklung	Informatik	12	deutsch
20	Management Science	BWL	12	englisch
20	Proceedings of the Annual Meeting of the ASIST	Infowiss.	12	englisch
24	IBM Systems Journal	Informatik	11	englisch

Tabelle 8: Zeitschriften die am öftesten zitiert wurden (mindestens 10 Zitate)

## 5 ISI auf dem Prüfstand

Wie bereits in der Einleitung erwähnt wurde, handelt es sich beim *Internationalen Symposium für Informationswissenschaft* um die bedeutendste *informationswissenschaftliche Veranstaltung im deutschsprachigen Raum*. Aufgrund des Anspruchs des Symposiums sollen nun folgende Punkte kritisch hinterfragt werden:

- Wie international ist die Tagung tatsächlich?
- Wie hoch sind wissenschaftliche Orientierung und Qualität?
- In welchem Maße hat das Symposium eine identitätsstiftende Wirkung für die deutschsprachige Informationswissenschaft?

### 5.1 Internationalität

Der Anspruch auf Internationalität kann in punkto Autorenschaft (und Publikationssprache) primär nur für die deutschsprachigen Länder erhoben werden. Wie die Autorenanalyse zeigte, kommen vier von fünf Autoren aus dem deutschen Sprachraum, zwei Drittel aller Autoren sind aus Deutschland. Dies deckt sich auch mit dem Sachverhalt, dass 81% aller Artikel in Deutsch verfasst wurden.

Andererseits ergab die Referenzanalyse, dass die Autoren durchaus internationale Forschungsergebnisse aufgreifen. So verweisen immerhin 57% der Referenzen auf englischsprachige Quellen. Auch im Ranking der am öftesten zitierten Zeitschriften waren großteils englischsprachige Zeitschriften enthalten. Wie in [Schlögl & Stock 04] nachzulesen ist, findet aber in umgekehrter Richtung so gut wie kein Informationsfluss statt. Dies trifft auch auf die beiden letzten Tagungsbände (ISI 2002 und 2004) zu, die in jener Publikation nicht berücksichtigt waren. Wie eine Recherche im Social Science Citation Index (Recherchedatum: 6. April 2007) ergab, wurden von den zwei Tagungsbänden insgesamt nur zwei Beiträge zitiert. In einem Fall handelte es sich um ein Selbstzitat, das zweite Zitat stammt ebenfalls von einem deutschsprachigen Autor, der zuvor am Internationalen Symposium für Informationswissenschaft teilgenommen hat.

### 5.2 Wissenschaftliche Orientierung

Bei der Betrachtung der Autoren handelt es sich eindeutig um eine wissenschaftliche Tagung. Im Durchschnitt waren drei Viertel aller Autoren in Forschung und Lehre tätig. Dies bestätigt auch die Institutionenanalyse, wonach 57% der Autoren von Universitäten waren. Relativ viele Beiträge wurden auch von Mitarbeitern von Forschungsgesellschaften oder ähnlichen Einrichtungen (zum Beispiel IZ Sozialwis-

senschaften, Gesellschaft für Mathematik und Datenverarbeitung, Fraunhofer Institut) verfasst.

Die Einschätzung des wissenschaftlichen Niveaus der Tagungsbände ist von den in diesem Beitrag berücksichtigten szientometrischen Indikatoren am ehesten mit der durchschnittlichen Länge der Literaturliste und der Halbwertszeit der Referenzen möglich. Bei beiden Merkmalen sind die Ausprägungen der internationalen informationswissenschaftlichen Spitzenzeitschriften zwei- bis dreimal höher.

Der Impact der ISI-Tagungsbände lässt sich mit den in diesem Beitrag durchgeführten Analysen nicht bestimmen. Wie aber die Grazia Colonia-Studie zeigte, werden die ISI-Tagungsbände kaum zitiert. Dies führte sogar zum paradoxen Ergebnis, dass von den zehn analysierten deutschsprachigen BID-Zeitschriften für die ISI-Proceedings der niedrigste Impact Factor berechnet wurde (siehe [Schlögl & Stock 04: 1158]). Dies liegt teilweise auch darin begründet, dass Tagungsbände mit Zeitschriften nur eingeschränkt vergleichbar sind. So haben Proceedings in der Regel geringere Auflagen als Zeitschriften und sind zudem oft nach dem Tagungsende vergriffen. Dies würde auch erklären, warum die ASIST Proceedings keinen nennenswerten Impact haben. Dass die Proceedings der letzten Symposien digital frei zugänglich sind, ist daher ein Schritt in die richtige Richtung. Um den Impact zu erhöhen und so einen entsprechenden Wissenstransfer gewährleisten zu können, müsste aber auch noch der Bekanntheitsgrad der Tagungsreihe außerhalb der „Stammkundschaft“ erhöht werden.

### **5.3 Identitätsstiftende Wirkung**

Unter diesem Punkt soll abschließend diskutiert werden, in wie weit ISI tatsächlich für sich beanspruchen kann, die deutschsprachige Informationswissenschaft zusammenzuführen. Wie das Institutionen-Ranking zeigte, sind vor allem solche Universitäten im Top-Ranking zu finden (Konstanz, Graz, Regensburg, Hildesheim, Saarbrücken), an denen informationswissenschaftliche Universitätsinstitute eingerichtet sind. Bis auf Hildesheim wurde an jeder dieser Universitäten bereits ein Symposium organisiert. Leider nicht gelungen ist es in der Vergangenheit, die restlichen Universitätsinstitute mit einem informationswissenschaftlichen Schwerpunkt (stärker) in die Tagung einzubinden.

Besonders gering war der Anteil von Autoren aus Fachhochschulen. Während 57% aller Autoren einer Universität angehörten, ist der entsprechende Anteil der FH-Autoren mit 6% sehr klein. Die Autorenanalyse brachte schließlich auch noch zum

Vorschein, dass sich die „Stammautoren“ aus einem relativ kleinen Kreis zusammensetzen.

Im Bereich der Referenzanalyse bringt die Zeitschriften-Selbstzitationsrate die Verbundenheit mit einem Periodikum am besten zum Ausdruck. Beispielsweise haben Verbandszeitschriften wie „Buch und Bibliothek“ oder „Library Journal“ Selbstreferenzraten von über 20% (siehe [Schlögl & Stock 04 : 1159]). Diese ist bei den ISI-Tagungsbänden mit 0,9% hingegen extrem niedrig. Es muss allerdings auch berücksichtigt werden, dass die Tagung seit 1992 nur mehr in zweijährigen Abständen stattfindet, was höheren Selbstzitationsraten entgegenwirkt.

## 6 Literaturverzeichnis

- [Albrecht 99] Albrecht, T. „Informationen für Studium und Beruf: Zitierpräferenzen von Verfassern betriebswirtschaftlicher Fachartikel“. In: WiSt (6) (1996), 319-324.
- [Baumgartner 05] Baumgartner, M. Eine szientometrische Analyse der Tagungsbände des Internationalen Symposiums für Informationswissenschaft. Dipl.-Arb. Universität Graz, 2005.
- [Dresel 03] Dresel, R. nfd. Information – Wissenschaft und Praxis 1981 – 2000. Eine informationswissenschaftliche Zeitschrift in szientometrischer Analyse. Dipl.-Arb. FH Köln, 2003.
- [Grazia Colonia 02a] Grazia Colonia. Informationswissenschaftliche Zeitschriften in szientometrischer Analyse. – Köln: FH Köln. – (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft, 33), 2003.
- [Grazia Colonia 02b] Grazia Colonia. „Zeitschriften des Bibliotheks- und Informationswesens im Vergleich“. In: Password (9) (2002), 23-28.
- [HI 06] Hochschulverband Informationswissenschaft. Internationales Symposium für Informationswissenschaft – ISI. <http://www.informationswissenschaft.org/> [17.1.2007]
- [Juchem 02] Juchem, K. „Der Bibliotheksdienst in szientometrischer Analyse“. In: Bibliotheksdienst 36 (2002), 1732-1740.
- [Juchem 04] Juchem, K. BuB: eine szientometrische Analyse im Zeitraum 1990 – 2003. – Dipl.-Arb. FH Köln, 2004.
- [Juchem, Schlögl & Stock 06] Juchem, K.; Schlögl, Ch. & Stock, W. G. „Dimensionen der Zeitschriftenszientometrie am Beispiel von „Buch und Bibliothek““. In: Information – Wissenschaft und Praxis 57 (2006), 31-37.
- [Koehler et al. 01] Koehler, W., et al. A profile statistics of journal articles: Fifty years of American Documentation and the Journal of the American Society for Information Science“. In: Cybermetrics, 4(1) (2000), online available at: <http://www.cindoc.csic.es/cybermetrics/articles/v4i1p3.html> [17.1.2007]
- [Knorz & Otto 00] Knorz, G. & Otto, Ch. Informations- und Wissensmanagement. Tagungsband für das Anwendungsorientierte Programm im Rahmen des 7.

- Internationalen Symposiums für Informationswissenschaft. Darmstadt: FH Darmstadt, Fachbereich Informations- und Wissensmanagement, 2000.
- [Resch & Schlögl 04] Resch, A.; Schlögl, Ch. „Die Wirtschaftsinformatik aus der Sicht ihres Hauptpublikationsorgans: Eine szientometrische Analyse der Zeitschrift „Wirtschaftsinformatik“/„Angewandte Informatik““. In: Wirtschaftsinformatik, 46(4) (2004), 302-310.
- [Schlögl & Stock 04] Schlögl, Ch. & Stock, W.G. „Impact and Relevance of LIS Journals: A Scientometric Analysis of International and German-Language LIS Journals – Citation Analysis versus Reader Survey“. In: Journal of the American Society for Information Science and Technology 55, (2004) 1155-1168.
- [Schlögl & Stock 06] Schlögl, Ch. & Stock, W. G. „Deutsche Zeitschriften des Bibliotheks- und Informationswesens: Leser, Zitate und Redaktionen in szientometrischer Analyse“. In: Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB) 53(5) (2006), 244-255.
- [Schubert 01] Schubert, A. „Scientometrics: the research field and it's journal“. In: Heck, A. (Hg.): Organizations and Strategies in Astronomy II. Dordrecht: Kluwer, 2001, 179-195.
- [Stock 01] Stock, W.G. „JCR on the Web. Journal Citation Reports: Ein Impact Factor für Bibliotheken, Verlage und Autoren?“ In: Password (5) (2001), 24-39.
- [Van Raan 97] Van Raan, A.F.J. „Scientometrics: State-of-the-Art“. In: Scientometrics, 38(1) (1997), 205-218.
- [White & McCain 89] White, H. D. & Mc Cain, K. W. „Bibliometrics“. In: Annual Review of Information Science and Technology, 24 (1989), 119-186.





# **Meta-Strategies to foster a Meta-Competence? Supporting information literacy at the workplace\***

*Susanne Mühlbacher*

Information Science  
University of Regensburg  
Universitätsstr. 31  
93051 Regensburg  
*susanne1.muehlbacher@sprachlit.uni-r.de*

## **Abstract**

Modern information workers need information-related competencies. Additionally, they have to be supported – technologically and conceptually – by an information landscape appropriate to their information-based tasks and needs. This thesis paper introduces a workplace-specific model of the information process, its related competencies and its main influential factors, based on a literary review. Afterwards, first findings of an empirical study, analyzing the scientific information process at a major pharmaceutical business organization, are presented. Subsequently, strategies to foster the information process will be discussed.

## **I Information Overload – Information Literacy**

Today's workplace demands a new kind of worker. Data is dispatched in picoseconds and gigabits and this deluge of information must be sorted, evaluated and applied. It is estimated that the average person spends 150 hours per year looking for information. (Alex Bennet, Information Literacy: A New Basic Competency. In CHIPS, Fall 2001)

The problem of human information processing and managing has been referred to in literature since the 1960s (Edmunds and Morris 2001). Much investigation has been conducted since in order to find applicable solutions, some of which have already been successfully introduced. Despite of this, the problem seems insoluble as scientific discourse continuously asks for innovative concepts and literacies: Our

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 61-75.

work and private life experience a growing overload of information that is characterized by a short lifecycle and an increasingly uncertain quality. Simultaneously, we are confronted with an increasing amount of technical tools to handle information. Information is nevertheless seen as the key for the success of governments, organizations as well as personal progress. Consequently, the required abilities of the information worker are constantly becoming more complex in our technology-based information society (Virkus 2003: 53, Kuhlen 1999, Edmunds and Morris 2001, Abell 2001).

There have been various efforts to define and teach these essential information handling abilities since the 1960s. Academic educators, information scientists and librarians, in the U.S., Australia and Europe, have conducted extensive research on information competencies and information management skills (Rader 1996-2000, Eisenberg 2004, Kuhlen 1999, Homann 1999). One outstanding concept is the information literacy movement that aims to enable the information worker to know “when and why you need information, where to find it, and how to evaluate, use and communicate it in an ethical manner” (CILIP in Ingold 2005: 15). This movement has often been referred to as a breakthrough within this field, attempting to combine the different approaches concerning information handling (Virkus 2003: 45). Various associations (AACRL, AASL, CILIP, NFIL, CILA, and SCONUL) have been formed with a common concern to find an appropriate, standardised definition of the concept and formulate information literacy competence standards. They developed learning objectives and information literacy curricula, embedding them in the general school and university curriculum at all levels. Starting in the U.S., this movement has “officially” just now reached schools and libraries in Europe, melting with the prevailing European concepts and initiatives (Virkus 2003: 45, Kuhlen 1999, Homann 1996/2002).

## **2 Workplace Information Literacy – a Problematic Transfer**

Within the field of industry and commerce, information literacy is stated to be a comparatively new concept (Donnelly and Craddock 2002, Lloyd 2004). As a means to efficiently conquer the growing amount of information and knowledge, business organizations eagerly integrated the techniques and tools of “knowledge management”. Nowadays, they are realizing that without appropriate educated information workers “the effort can be wasted and costly mistakes made [by] adopting the practices without understanding the principles” (Houghton and Halbwirth 2002; cf. Kuhlen 1999). Lately, the field of industry and commerce shows a rising

interest in topics as usability, Web 2.0 applications and “information literacy” or rather “the people component of knowledge management” (Oman 2001) – as a means to integrate the so far forgotten user-oriented aspect into the business information landscape (Lehner 2000, Houghton and Halbwirth 2002, O’Sullivan 2002).

Current models of teaching curricula of information literacy are, however, mostly restricted to the mere implementation, analysis and teaching of the usage of the established technical applications. This shows the prevailing ambiguity and the need for an appropriate definition of the concept within the workplace environment: Information literacy is often confused with skill-based literacies as computer literacy, library literacy or research skills. What is more, information literacy models often integrate various well-used and established concepts as time management, information management, teamwork, data mining, or internet-experience (...). (O’Sullivan 2002: 10/11) Thus, companies do not recognize the difference of information literacy to these well-established soft skills. Besides, companies seem to take it more or less as granted that their employees have acquired certain basic literacy skills in higher education, not realizing that a transfer to the constantly changing workplace environment may be difficult or impossible (Lloyd 2004, Cheuk 2002). Furthermore, educational information literacy models are far too strategic concepts created and used by librarians and information specialists, with the aim to standardize required information skills. Thus, the transfer to the workplace context is problematic and an appropriate definition – let alone a curriculum for working professionals and concepts for its sensible integration into the business landscape – is still missing (Tuominen 2005: 330, Cheuk 2002).

### **3 Workplace Information Literacy – Current Definitions**

Most prevailing definitions of workplace information literacy restrict it to a mere set of basic abilities of obtaining and using information effectively (Goad 1999, Abell 2001, Cheuk 2002):

[Information literacy embodies] a set of abilities for employees to recognize when information is needed and to locate, evaluate, organize and use information effectively, as well as the ability to create, package and present information effectively to the intended audience. (Cheuk 2002:2)

Additionally, the information literate individual is claimed to own a variety of meta-competencies, or rather a set of higher or related abilities. These meta-skills consist of personal and social competencies as well as a basic knowledge of the thematic environment (Mötsch 1997 in Kuhlen 1999, Stock 1998 in Kuhlen 1999).

According to Bruce (1999) an information literate person has developed an information style, uses information technology and systems, has knowledge of the world of information and owns values which promote information use.

The combination of this set of meta-skills and basic skills enables an information literate person to approach information critically, implement information processes, conquer the information chaos, analyze and solve complex problems, create new meanings and share knowledge locally and globally (Kuhltau 1993 in Tuominen 2005: 335, Bundy 2004: 4, Bruce 1999, Kuhlen 1999, Stock 1998 in Kuhlen 1999, Tuominen 2005, Lloyd 2004). Thus, it enables an information literate person to become an independent, self-directed and self-sufficient learner, able to “make [his/her] [...] world rather than conforming [his/her] [...] world made by others” (Williams 2003 in Tuominen 2005: 335, Kuhlen 1999).

Webber and Johnston (2003) summaries the described characteristics in a general definition of workplace information literacy:

The adoption of appropriate information behaviour to obtain, through whatever channel or medium, information well fitted to information needs, together with a critical awareness of the importance of wise and ethical use of information in society. (Webber 2003: 6)

Their definition incorporates the main actors that coin information literacy:

<p><b>INFORMATION WORKER</b> The human and his/her individual information behaviour depending on his information needs due to task and environment.</p>	
<p><b>INFORMATION SOURCE</b> the information source and its variable character (whatever channel or medium).</p>	<p><b>INFORMATION SOCIETY</b> The society and communication with society within existing values and borders.</p>

The major influencing environmental components are, in their eyes, “information economy“, “organizational culture“, “specialist subject field/technical progress“ as well as “personal, social and cultural factors”. These factors can change over time and generate individual forms of information literacy (Webber and Johnston 2000: 385).

## 4 Workplace Information Literacy – a Dynamic Model

On basis on the literary review (see 3) one can assign four major dimensions that influence the information process and therefore information literacy: the basic skills

dimension (instrumental/basic skills), the dimension of the individual worker (personal/social competencies), the business organization dimension (environmental conditions) as well as the society and its changeability.

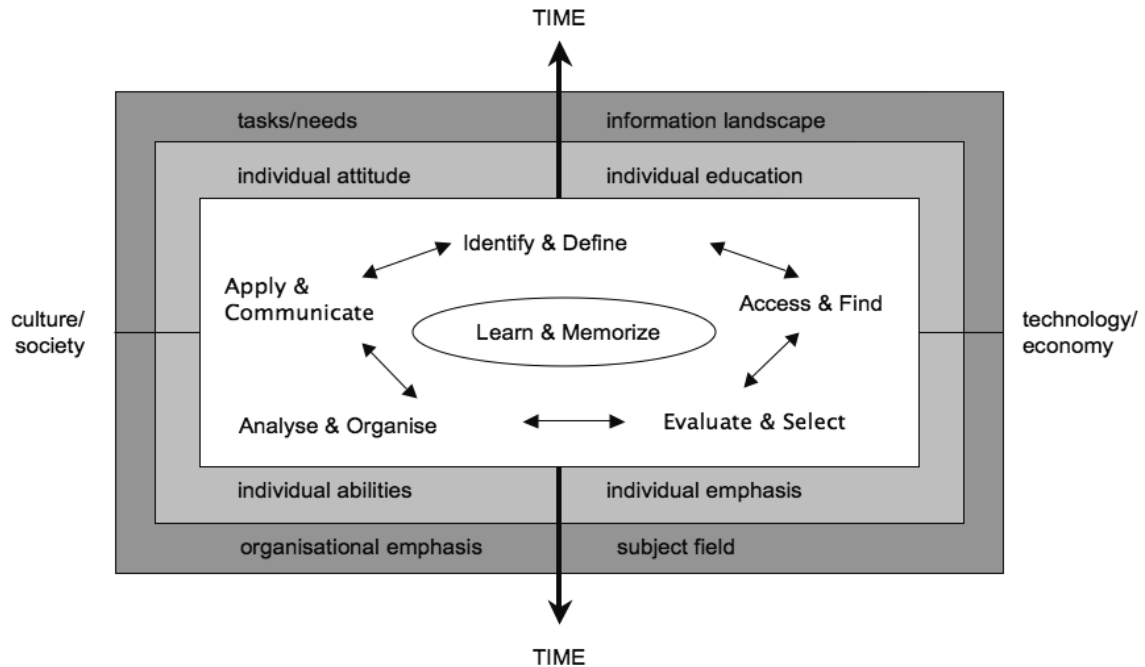


Figure 1: A dynamic model of information literacy and its main influential factors

The six steps of the basic skills dimension (white square in the center) are based on the established information literacy standards (Goad 1999, ALA 1989, Armstrong 2005, SCONUL 1999).

Step	Definition
Identify & Define	<ul style="list-style-type: none"> <li>• Define an information need</li> <li>• Form the information need into a research question</li> <li>• Break the research subject down into parts</li> <li>• Analyze the relationships and hierarchies</li> </ul>
Find & Access	<ul style="list-style-type: none"> <li>• Develop a question list</li> <li>• Select multiple potential information sources (human, technical, physical)</li> <li>• Formulate adequate search strategies</li> <li>• Search effectively within information sources</li> </ul>
Evaluate & Select	<ul style="list-style-type: none"> <li>• Filter information while remaining focused</li> <li>• Evaluate quality of information</li> <li>• Select information appropriate to need</li> <li>• Review the research process critically</li> </ul>
Analyse & Organise	<ul style="list-style-type: none"> <li>• Interpret, analyse and synthesise information</li> <li>• Manage and classify information</li> <li>• Process and manipulate information</li> <li>• Record and store information</li> </ul>

Apply & Communicate	<ul style="list-style-type: none"> <li>• Put information in a context</li> <li>• Apply information for problem-solving</li> <li>• Present and communicate information to adequate authorities</li> <li>• Communicate information considering economic, cultural, social, ethical and legal issues</li> </ul>
Learn & Memorize	<ul style="list-style-type: none"> <li>• Continuously incorporate new knowledge into knowledge base</li> <li>• Contribute actively to the creation of new knowledge</li> <li>• Actively update knowledge regularly</li> <li>• Access and apply knowledge according to information need if necessary</li> </ul>

The order of the steps of the information process is not necessarily linear. People may e. g. refine the definition of their information needs after having accessed relevant information sources and having evaluated and selected relevant information. Just, as they may reorganise their information after its application and communication.

The “individual” dimension (light grey square) incorporates four components which all have a major influence on the carrying out of the process.

<i>Component</i>	<i>Definition</i>
Individual attitude	Positive or negative feelings that people associate with the information process, e. g. regarding quality, benefit, expenditure and the task itself
Individual education	Level of experience and training of people regarding the basic skills of the information process and related tools (search courses, tool courses, etc.) as well as thematic background and education.
Individual abilities	Individual abilities and meta-competencies people own apart from the basic information skills. This includes e. g. critical awareness, level of knowing the information universe, problem-solving, decision-making, social responsibility, ability of communication, search skills, higher order analysis, organisational skills, creativity (etc.)
Individual emphasis	Completeness and importance of the steps that people individually associate with the information process. This is closely related to their tasks as well as to their meta-competencies.

The “organisational” dimension (dark grey square) incorporates the given tasks/needs, the subject field, the organisational focus regarding the process steps and the provided information landscape.

<i>Component</i>	<i>Definition</i>
Tasks/Needs	The needs of the business organisations and the related tasks of the information worker.
Subject field	The thematic field in which the information process is carried out.
Organisational emphasis	Completeness and promoted importance of steps that the business organisation associates with the information process and therefore supports technologically, organisationally and psychologically.

Information landscape	Supportive technical devices and applications (information sources, etc.), social interaction (team, colleagues, personal network, etc.), time & money resources (workplace conditions) provided by the business organisation within the limits of the technical development and social and financial barriers.
-----------------------	---

The society dimension (surrounding white area) embodies the coining and adaptation of the process and its related skills by economic, social, cultural and technological conditions and changes due to time (time axis).

The level and requirements of information literacy are therefore dependent on the characteristics of the different dimensions and the environmental influences as e. g.:

*Organisation:* The model – and therefore the requirements – for employees working at a law firm may differ from employees working in a car company regarding subject field, organisation support/emphasis and information landscape, basic skills (etc.).

*Task:* The model for a secretary and a lawyer working at the same company may differ e. g. regarding attitude, educational background, tasks/needs, basic skills and provided information landscape (etc.).

*Society:* The model for a secretary in Germany may differ from that of a secretary in India regarding attitude, basic skills (etc.).

In conclusion, the model is not to be used to test or classify information workers concerning their level of information literate competence. It is rather a means of a context-, human-, and time- sensitive analysis of the information process and individual strategies to support information-based workflows.

## **5 Workplace Information Literacy – a Concept of Analysis**

Attempts to analyse information literacy at the workplace are mostly based on interviews. Researches are rarely mentioning the employment of standardised methods, empirical proof or methods of analysis. Furthermore, established standards how to analyse the requirements and problems of information literacy at the workplace do not yet exist. Oman (2001) gives a rather informal recommendation: She suggests assessing the information infrastructure, the demographics of employees, the information process, and individual information literacy competencies. Additionally, she strengthens the great importance to define what information literacy means for the individual organization, workplace and tasks.

In the course of a doctoral thesis an analytical and empirical case study is conducted at a major pharmaceutical company. The aim is to define problematic areas and general applicable strategies to analyse and advance the scientific information process in business organisations. The workplace information analysis is carried out according to the introduced model of information literacy.

*Information process – basic skills*

- Step 1: Identify & Define
- Step 2: Find & Access
- Step 3: Evaluate & Select
- Step 4: Analyse & Organise
- Step 5: Apply & Communicate

*Information worker*

- Attitude
- Individual emphasis
- Education
- Meta-competencies

*Business organisation*

- Organisational emphasis
- Subject/Field
- Information landscape
- Tasks & Needs

*Information Society*

- Cultural aspects and changes
- Technological aspects and changes
- Economical aspects and changes

The four dimensions are being investigated from three different point-of-views in order to prevent a singular view on the situation:

	Analysis	Information process	Information worker	Business organisation	Information Society
Method					
Expert interviews (information professionals)	X	X	X	X	X
Information worker interviews	X	X	X	X	X
Independent expert analysis	X	X	X	X	X

## 6 Workplace Information Literacy – a Case Study<sup>1</sup>

### 6.1 Interviewing Information professionals – First Findings

The first part of the analysis in the course of the doctoral thesis consists of unstructured information professional interviews. Participants are eight employees of the scientific information department of a major pharmaceutical organisation. The scientific information department consists of the scientific research partners, the library and the intranet and scientific database unit. They are responsible for the provision, research and management of scientific information and work closely together with the scientists. The dynamic model of the information process was shown to the participants as a basis of the discussion. One interview lasted approx. 45 Minutes.

<sup>1</sup> This is only a small part of the findings of the empirical study as the study is still in progress.



The identified problems regarding the information worker, the business organisation and information landscape are presented in relation to recent studies regarding problematic areas of information literacy.

	<i>Problematic Areas – information process (basic skills)</i>	<i>Cheuk</i> <sup>2</sup>	<i>D&amp;J</i> <sup>3</sup>	<i>W&amp;J</i> <sup>4</sup>	<i>M</i> <sup>5</sup>
1	Unable to determine the nature and extent of information needed	X			
2	Unable to retrieve effectively from information systems due to a lack of basic information skills and strategies	X	X	X	X
3	Difficulties in locating information/ lack of knowledge of range of internal databases and external information sources available/only using established standard tools	X		X	X
4	Unable to evaluate and filter information	X			
5	Unable to manage the information and email overload	X			
6	Unable to exploit technology to manage information	X			
7	Unable to relate information creation and use to a broader context	X			
8	Unethical use of information	X			
9	Unable to evaluate the costs and benefits of information management	X			
10	Lack of awareness of the importance of information literacy	X			X
11	Difficult communication between information research professionals and knowledge workers				X

	<i>Problematic Areas – business organisation</i>	<i>Cheuk</i>	<i>D&amp;J</i>	<i>W&amp;J</i>	<i>M</i>
1	Unable to recognize information needs		X		
2	Unable to locate and access information		X		
3	Unable to organise, apply and communicate		X		X
4	Unable to synthesis and create/No clearly organised information landscape		X		X
5	Lack of comprehensive training programme	X	X	X	X
6	Lack of awareness of importance of information literacy		X	X	X
7	No clearly organised information landscape (e. g. Intranet)				X
8	Difficult handling of Information search tools				X
9	Growing amount of information sources complicate adequate selection and usage				X

<sup>2</sup> Cheuk (2002: 3-5).

<sup>3</sup> Donnelly and Craddock (2002: 2).

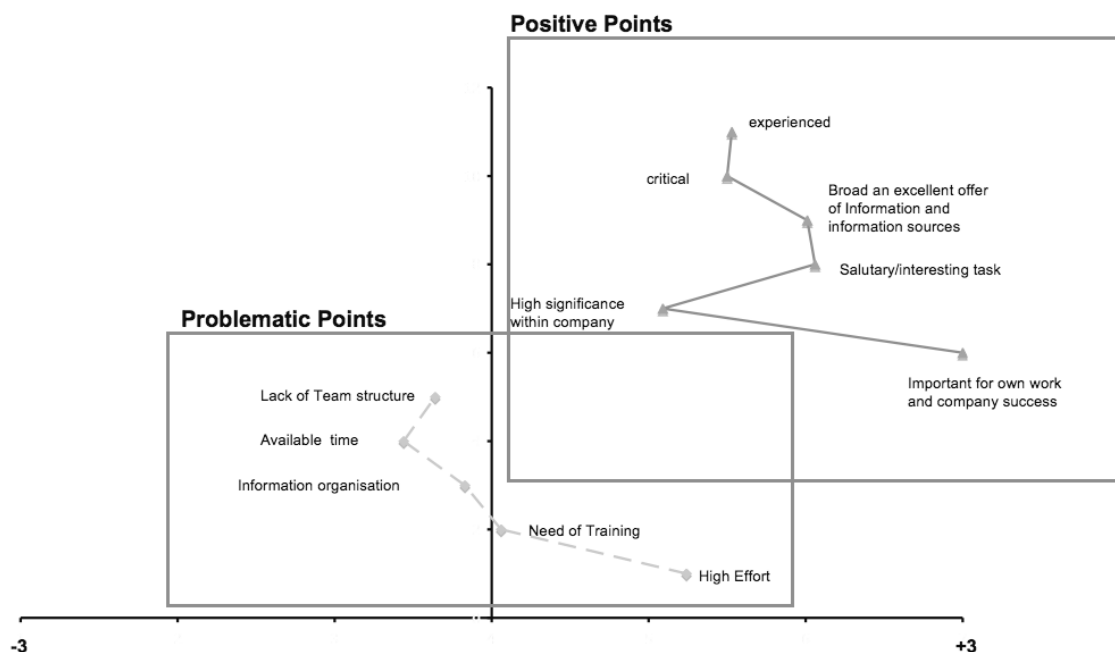
<sup>4</sup> Webber and Johnston (2001: 12).

<sup>5</sup> Mühlbacher expert interviews 2006.

## 6.2 Interviewing Information Workers – First Findings

The second part of the analysis consisted of structured information worker interviews. The aim was to identify the main problematic areas participants experience in the course of the information process. Participants were 78 scientists working in the research and development unit of a major pharmaceutical company. The interview took approx. one hour.

The participants show a very positive attitude towards the scientific information process and their own competencies regarding this process. They see themselves as experienced and critical information workers. Furthermore, they experience scientific information work as an interesting and informative task which broadens their mind and their knowledge. The company itself attaches, in their eyes great importance on scientific information work. This is supported by the broad and excellent offer of information and information sources within the business information landscape.



Nevertheless, the participants identified some major problematic areas within their information work. They especially criticised the current conditions for team working within this process. They experience a growing development towards solitary work. One reason for this is the lack of organised and sensible platforms to communicate, especially internal knowledge. Current technologies do, in their eyes, not fit the task or are neglected for the high effort of maintenance.

Due to the technological development more and more information sources become available. Consequently, it gets more difficult to choose the best source for one's task. Participants state that they usually rely on 2-3 information sources. However,

they do have little time and interest to work with new applications. As there do not exist common standards for search possibilities within information sources, they fear that it costs too much effort to learn the handling of new applications. Additionally, they state that the growing amount of available information leads to an information overload and chaos, making it difficult to evaluate and select reliable information. Equally, there seems to be little time, but great need for further training regarding information skills. Because of the shifting of administrative and bureaucratic tasks from responsible departments to the individual workplaces, their time for research and the scientific information process shrinks constantly. Furthermore, the scientific information process requires high effort to guarantee its quality. Keeping up-to-date regarding current developments in literature shifts continuously into a weekend-hobby.

## **7 Strategies to Support Workplace Information Literacy**

Supporting information literacy is not a simple task. On the one hand, the information worker is continuously in need of training and adapting his skills to the changing information landscape. This affords a comprehensive, ongoing and user-oriented training curriculum, integrated in the business landscape. Furthermore, new information sources and technologies and their benefits are to be marketed via intranet and seminars in order to keep the information worker up-to-date. On the other hand, the information workers as well as the companies have to be made aware of the necessity of information skills. This requires the promotion of the importance of the task as a critical business skill. In order to activate the information worker advance his information skills, it is recommended giving tangible reward as an encouragement to acquire these skills (see Cheuk 2002: 8; Donnelly and Craddock 2002). Thus, the basic strategies to support information literacy are:

<i>Advancing information literacy at the workplace – Step 1</i>	
Information worker/Information Process	Training
	Promotion
	Reward

However, as information literacy is a meta-competence, influenced by various factors, these three steps may be important but are certainly not sufficient. In order to advance the information skills at the workplace more effort is needed.

First, the information landscape needs to be adapted to the user needs: This means clearly organised platforms (as e. g. the intranet), easy accessible and usable standardized tools and a greater support of internal communication.

<i>Advancing information literacy at the workplace – Step 2</i>	
Business Organisation/Information landscape	Clearly organised information landscape
	Easy-to-use and easy-accessible information sources
	Supporting communication and organisation of internal knowledge

Second, fears and prejudices regarding the communication of information need to be dissolved. This may be a long-lasting process, which should actually be started at school. Additionally, time schedules should be relaxed to ensure the quality of the information process, that requires high effort and therefore time. Last but not least, administrative tasks should be directed back to the responsible departments; scientists should stay scientists rather than becoming clerks. It is a fact that Bureaucracy is a cultural problem disabling the creativity and innovative ability (see Grether 2005).

<i>Advancing information literacy at the workplace – Step 3</i>	
Information society/Cultural aspects	Dissolving of prejudices against sharing of information
	Downsizing of Bureaucracy
	Speed culture versus Quality

## 8 Conclusion

Literary research as well as empirical study has shown that the concept of information literacy at the workplace slowly but surely takes shape. It crystallizes that a model for information literacy at the workplace has to be a meta-concept that unites different competencies and considerate various influential factors. Due to this flexible characteristic, methods of analysis as well as strategies to support information literacy cannot be singular. A successful analysis of information literacy at the workplace requires a definition of a workplace-specific model of information literacy. This model should be based on the analysis of four dimensions: the information process, the information worker, the business organisation and the information society. Furthermore, external influences changed by time as social and cultural development and conditions as well as technical and economical progress and conditions have to be considered. This model helps to detect deficiencies and problems without trying to classify human beings. On basis of the results of the described analysis, supporting meta-strategies can be established regarding the different dimensions. These meta-strategies include training and education of the information worker, promotion of the importance of information work for personal and company success by appropriate promotion and incentives. Furthermore, the informa-

tion landscape has to be adapted to the user needs and user abilities. Last but not least, cultural, organisational and personal barriers, e. g. to share information, to give time to do a task (etc.) need to be identified and dissolved.

## 9 Bibliography

- Abell, A. ; Oxbrow, N. (2001). *Competing with knowledge*. London: Library Association Publishing.
- American Library Association Presidential Committee on Information Literacy (1989). Final Report, Chicago. <http://www.ala.org/ala/acrl/acrlpubs/whitepapers/presidential.htm> (03.08.06).
- Armstrong, C. (2005). Defining information literacy for the UK. *Library and Information Update*, Jg. 4, Nr. 1-2, S. 22-25.  
<http://www.cilip.org.uk/publications/updatemagazine/archive/archive2005/janfeb/armstrong.htm> (06.01.07)
- Bennet, Alex (2001). Information Literacy – A New Basic Competency. *CHIPS Fall 2001*.  
[http://www.chips.navy.mil/archives/01\\_fall/information\\_literacy.htm](http://www.chips.navy.mil/archives/01_fall/information_literacy.htm) (06.01.07)
- Bieler, D.; Hapke, T. (2005). Lernen, Informationskompetenz und Visualisierung – Das Online-Tutorial DISCUS (Developing Information Skills & Competence for University Students) der Universitätsbibliothek der TU Hamburg-Harburg. In: *ABI-Technik* 25 (3): 162-181.
- Bundy, Alan (Hg.) (2002). *Australian and New Zealand information literacy framework. Principles, standards and practice*, 2. Auflage, Australian and New Zealand Institute for Information Literacy: Adelaide. <http://www.anu.edu.au/caul/info-literacy/InfoLiteracyFramework.pdf>
- Bruce, C. (1999). Workplace experiences of information literacy, *International Journal of Information Management* 19, p33-47.
- Cheuk, B. W. (2000). Exploring information literacy in the workplace: a process approach, In: Bruce, C.S. and Candy, P.C. *Information literacy around the world: advances in programs and research*. Wagga Wagga, NSW: Charles Sturt University. pp. 177-191.
- Cheuk, B. W. (2002). *Information literacy in the workplace context: issues, best practices and challenges*. White Paper prepared for UNESCO, the U.S. National Commission on Libraries and Information Science, and the National Forum on Information Literacy, for use at the Information Literacy Meeting of Experts, Prague, The Czech Republic, July, 2002. <http://www.nclis.gov/libinter/infolitconf&meet/papers/cheuk-fullpaper.pdf> (09.12.06)
- Council of Australian University Librarians (2001). *Information Literacy Standards*. Canberra: Council of Australian University Librarians. <http://www.caul.edu.au/caul-doc/InfoLitStandards2001.doc> (09.12.06)

- Donnelly, A; Craddock, C: (2002). Information literacy at Unilever R&D, *Library and Information Update* 1. <http://www.cilip.org.uk/update/issues/deco2/article2dec.html>. (22.01.07)
- Edmunds, A.; Morris; A. (2001). The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management* 20: p. 17-28.
- Eisenberg, M. B.; Spitzer, K. L.; Lowe, C. A. (2004). *Information literacy: essential skills for the information age*. 2nd edition. Syracuse: ERIC Clearinghouse on Information & Technology.
- Grether, T. (2005) *Weniger Bürokratie.in Deutschland wagen*. Ein Bericht der Bertelsmannstiftung 12/2005. [http://www.bertelsmannstiftung.de/bst/de/media/xcms\\_bst\\_dms\\_16398\\_\\_2.pdf](http://www.bertelsmannstiftung.de/bst/de/media/xcms_bst_dms_16398__2.pdf)
- Houghton, J.M.; Halbwirth, S.(2002). Knowledge management and information literacy: A new partnership in the workplace?, In: P.Ward (eds). *Continuing professional education for the information society: Proceedings of the Fifth World conference on Continuing Professional Education for the Library and Information Science Professions*. Saur: München. pp. 70 – 79:
- Homann, Benno (1996). Schulungen als Aufgabe einer Benutzer orientierten Bibliothek. *Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB)* 43(6), p 595-613.
- Homann, Benno (2002). Standards der Informationskompetenz. Eine Übersetzung der amerikanischen Standards der ACRL als argumentative Hilfe zur Realisierung der "Teaching Library. *BD* 36 (5); p. 625-637 .
- Ingold, M. (2005). *Informationskompetenz: ein (neues) Leitbild für betriebliche Informationsstellen?*. Leitbild Informationskompetenz: Positionen, Praxis, Perspektiven im Europäischen Wissensmarkt – proceedings der 57. Jahrestagung der DGI: 15-26.
- Kirk, J. (2004). *Information and work: extending the roles of information professionals, Challenging ideas*. Proceedings of the ALIA 2004 Biennial Conference, 21-24 September 2004, Gold Coast. <http://conferences.alia.org.au/alia2004/pdfs/kirk.j.paper.pdf>. (05.01.06)
- Klatt, R.; Gavrilidis, K. (2001). Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung Barrieren und Potenziale der innovativen Mediennutzung im Lernalltag der Hochschulen – Endbericht. Bundesministeriums für Bildung und Forschung.
- Krause, Ch.; Pearce, J. (1999). ORNLS: The First 50 Years – Chapter 5: Balancing Act. *Oak Ridge National Laboratory Review* 25 (3 and 4).
- Kuhlen, R. (1999). *Die Konsequenzen von Informationsassistenten*. Frankfurt a. Main: Suhrkamp Taschenbuch Verlag.
- Lazonder, A. W.; Biemans, J.A.(2000). Differences between Novice and Experienced Users in Searching Information on the World Wide Web. *Journal of the American Society for Information Science and Technology* 51 (6).
- Lehner Franz (2000). *Organisational Memory: Konzepte und Systeme für das organisatorische Lernen und das Wissensmanagement*. München: Hanser Verlag.
- Lewis, D.(1996). *Dying for Information?* London: Reuters Business Information. p.2.

- Lloyd, A.(2003). Information literacy: the meta-competency of the knowledge economy? An exploratory paper. *Journal of Librarianship and Information Science* 35(2), p 87-92.
- Lloyd, A. (2004). Working (in)formation: conceptualizing information literacy in the workplace, Lifelong learning: whose responsibility and what is your contribution? Refereed papers from the 3rd International Lifelong Learning Conference, Yeppoon, Queensland, Australia, 13-16 June 2004, Central Queensland University Press, Rockhampton, p 218-224.
- Oman, Julie (2001). Information literacy in the workplace. *Information Outlook* 5 (6), S. 32-43.
- O'Sullivan, C. (2002). Is information literacy relevant in the real world? *Reference Services Review* 30 (1), p7-14.
- Rader, Hannelore B. (1996-2000). Library instruction and information literacy. *Reference Services Review*, Jg. 24 – Jg. 28.
- SCONUL Advisory Committee on Information Literacy (1999). *Information skills in higher education: A SCONUL position paper*.  
[http://www.sconul.ac.uk/activities/inf\\_lit/papers/Seven\\_pillars2.pdf](http://www.sconul.ac.uk/activities/inf_lit/papers/Seven_pillars2.pdf) (07.12.06)
- Tuominen, K; Savolainen, R.; Talja, S. (2005). Information Literacy as a Sociotechnical Practice. *The Library Quarterly*, 75, p 329 – 345
- Virkus, Sirje (2003). Information literacy in Europe: a literature review. *Information Research*, 8(4), Art. 159. <http://informationr.net/ir/8-4/paper159.html> (28.10.2005)
- Webber, S. and Johnston, B. (2000). Conceptions of information literacy: new perspectives and implications. *Journal of Information Science* 26 (6), p381-397.
- Webber, S. and Johnston, B. (2003) Assessment for information literacy: vision and reality. In: Martin, A. and Rader, H. (Eds). *Information and literacy: enabling learning in the 21st century*. London: Facet Publishing. ISBN 1-85604-463-7. p 101-111





# Qualität der *Wikipedia*

## Eine vergleichende Studie\*

*Rainer Hammwöhner, Karl-Peter Fuchs,  
Markus Kattenbeck, Christian Sax*

Informationswissenschaft  
Philosophische Fakultät IV  
Sprach- und Literaturwissenschaften  
Universität Regensburg

### **Zusammenfassung**

Die Qualität steht von Beginn an im Vordergrund der Debatte um die *Wikipedia*. In diesem Beitrag werden zunächst Studien vorgestellt, welche die Qualität der *Wikipedia* zum Gegenstand haben, sodann werden Design und Ergebnisse einer Studie präsentiert, welche eine Kollektion von Qualitätsparametern im Vergleich zwischen der deutschen *Wikipedia* und dem *Großen Brockhaus* evaluiert.

### **I Einleitung**

Prinzip und Erfolg der *Wikipedia*<sup>1</sup> sind in mancherlei Hinsicht bemerkenswert. Zunächst fällt vermutlich der rein quantitative Aspekt ins Auge: Die englische *Wikipedia* [Wikipedia 2007] hat mittlerweile (Stand 6.1.2007) mehr als 1,5 Millionen Artikel, weitere 11 in anderen Sprachen verfasste *Wikipedias* verfügen über mehr als einhunderttausend Beiträge. Beeindruckend ist auch das Wachstum der *Wikipedias*, das exponentielle Wachstumsphasen einschließt [Voss 2005, S. 27 ff.]. Der eigentliche Grund für die Aufmerksamkeit, welche die *Wikipedia* genießt, liegt jedoch in dem augenscheinlichen Konflikt zwischen dem Anspruch der *Wikipedia*, enzyklopädisches Wissen ohne Einschränkung in Qualität und Reichweite bereitzustellen,

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 77-90.

<sup>1</sup> Einen guten Überblick über die Prinzipien der Wikipedia bietet [Danowski/Voss 2005].

auf der einen Seite und dem Prinzip, dieses Wissen von einer offenen, z. T. anonymen Teilnehmergruppe zusammen stellen zu lassen, auf der anderen. Auf diese Herausforderung für den traditionellen Wissens- und Wissenschaftsbetrieb mit seinen Verfahren der Qualitätssicherung erfolgten sowohl vorsichtig affirmative [Kuhlen 2005] als auch polemisch ablehnende Reaktionen [Lanier 2006], wobei letztere auch durch schwerwiegende Qualitätsprobleme in der *Wikipedia* motiviert wurden. Große Beachtung wurde einer Studie über die Qualität der *Wikipedia* zuteil, die im Auftrag von *Nature* erstellt wurde [Giles 2005]. Der *Wikipedia* wurde dort eine Qualität zugesprochen, welche jener der *Encyclopedia Britannica* gleich zu setzen sei. Diese Studie wurde von Seiten der *Britannica* angegriffen, von *Nature* aber verteidigt und mit Materialien untermauert<sup>2</sup>. Weniger als das Ergebnis dieser Debatte ist ihr Verlauf von Interesse. Es wurde deutlich, dass ein Konsens über eine angemessene Methodik zur Evaluierung der *Wikipedia* – wie auch ihrer Konkurrenzprodukte – erst noch zu erzielen ist. Hier ist ein erster Gegenstand für eine Erforschung der *Wikipedia* gegeben. Es überrascht also nicht, dass Fragen zur Qualität einen nennenswerten Umfang der bisher noch nicht sehr umfangreichen, aber durchaus qualitätsvollen Forschung zur *Wikipedia*<sup>3</sup> einnehmen.

## 2 Studien zur Qualität der *Wikipedia*

Zunächst sollen einige Studien vorgestellt werden, die sich mit Aspekten der Qualität der *Wikipedia* befassen. Diese werden jeweils hinsichtlich ihrer Methodik und Reichweite untersucht. Auf der Grundlage dieser Überlegungen wurde eine neue Studie entworfen, deren Ergebnisse in diesem Beitrag präsentiert werden.

### 2.1 Die Nature-Studie [Giles 2005]

Im Folgenden soll der Ablauf der Studie kurz skizziert werden. Zunächst wurden 50 Artikel ausgewählt, die ein breites naturwissenschaftliches Themenfeld abdecken sollten. Die Darstellungen aus der *Wikipedia* und der Online-Ausgabe der *Encyclopedia Britannica* sollten sich ungefähr im Umfang entsprechen, so dass letztendlich nur 42 Artikel zur Evaluation verblieben. Die Themen waren so ausgewählt, dass in der Redaktion Kompetenz vorhanden war. Zu jedem der Themen wurde ein Exper-

---

<sup>2</sup> Verweise auf diese Materialien finden sich in der Online-Fassung von [Giles 2005].

<sup>3</sup> Im Zusammenhang mit der deutschen Wikipedia wurde für *Wikipedia*-Forschung der Terminus *Wikipedistik* geprägt. Zusammenstellungen von Publikationen über die Wikipedia finden sich unter [Wikipedia 2007a, 2007b].

te ausgewählt, der jeweils ein Paar aus einem *Wikipedia*- und einem *Britannica*-Artikel zu bewerten hatte, ohne die Zuordnung zu kennen. Die Gutachter sollten in den Artikeln drei Arten von Fehlern identifizieren: sachliche Fehler, kritische Auslassungen, irreführende Formulierungen. Die Gutachten wurden dann vom Redaktionsteam gesichtet, die Fehlerangaben wurden gewichtet und summiert. Es stellte sich heraus, dass die beiden Enzyklopädien keine Unterschiede hinsichtlich gravierender Fehler aufwiesen (je 4), in Artikeln der *Wikipedia* aber deutlich mehr „kleine“ Fehler zu beobachten waren (162/123). Besagte Studie wurde durch die Redaktion der *Britannica* heftig kritisiert, da Ungereimtheiten hinsichtlich der Textgrundlage und der Bewertung von Fehlern bestünden. Die Kritik wurde von *Nature* zurückgewiesen.

Unabhängig von dieser Kritik ist jedoch zu beachten, dass die Reichweite der Studie sehr begrenzt ist. Folgende Einschränkungen sind zu bemerken:

- Die Evaluation evaluiert nur die Korrektheit von Lexikoneinträgen. Andere Qualitätsaspekte wie Lesbarkeit und Verständlichkeit, Konsistenz in der Gliederung, thematische Abdeckung etc. blieben unberücksichtigt. Die Vermutung, dass sich die Qualität redaktioneller Arbeit gerade bei diesen Qualitätsaspekten zeigen kann, ist nahe liegend.
- Die ausgewählten Artikel repräsentieren nicht die zu untersuchenden Grundgesamtheiten (Artikel der *Wikipedia* bzw. *Britannica*). Es ist davon auszugehen, dass naturwissenschaftliche Fragestellungen andere *Wikipedia*-Autoren anziehen als etwa Themen der Allgemeinbildung. Die Studie besitzt also Aussagekraft allenfalls für naturwissenschaftliche Themen.

## 2.2 Studie von Christian Schlieker [Schlieker 05]

In seiner Diplomarbeit befasst sich Herr Schlieker mit elektronischen Enzyklopädien. Neben anderen Fragestellungen, die hier nicht weiter erörtert werden sollen, nimmt eine komparative Studie bezüglich der Qualitätsparameter Abdeckung und informationelle Absicherung breiten Raum ein. Diese Studie ist vor allem deshalb von Interesse, weil sie methodisch und inhaltlich gut vergleichbar mit der in diesem Artikel neu vorgestellten Studie ist.

Untersucht werden die deutsche *Wikipedia*, *Encarta* und eine Druckfassung der Enzyklopädie *Der Große Brockhaus* (20. Auflage). Aus jeder der Enzyklopädien wird eine Zufallsstichprobe von 15 Artikeln gezogen [Schlieker 2005, S 43 ff]. Dabei werden nur Vollartikel und keine Begriffsklärungen o. Ä. berücksichtigt. Aus der *Wikipedia* wird die Stichprobe mit Hilfe der Zufallsfunktion gezogen, aus der

*Encarta* mangels besserer Alternative durch „zufälliges“ Scrollen im alphabetischen Index bestimmt und aus der *Brockhaus*-Enzyklopädie durch ein mehrstufiges Zufallsverfahren (random walk), das zunächst den Band, dann Spalte, Seite und Lemma auswählt. Durch das Auswahlverfahren dürften aus dem *Brockhaus* mit leichter Präferenz längere Artikel ausgewählt werden, während bei der *Encarta* keine gleichmäßige Berücksichtigung des Alphabets zu erwarten ist. Kritischer für die Aussagekraft der Untersuchung dürfte jedoch der geringe Stichprobenumfang sein. Die Stichproben benutzt Herr Schlieker [Schlieker 2005, S. 51 ff] zunächst, um die thematische Abdeckung der Enzyklopädien zu untersuchen:

- Er überprüft in welchem Maße die einzelnen Enzyklopädien die Gesamtmenge von 45 Lemmata abdecken. Es ergibt sich die Rangfolge Brockhaus – Wikipedia – Encarta (33, 29, 25 von 45).
- Es wird versucht, die 45 Lemmata zu kategorisieren und dann aus der jeweiligen Abdeckung auf thematisch Profile der Enzyklopädien zu schließen. Brockhaus wird ein geringerer Aktualitätsbezug unterstellt als der Wikipedia, während bei der Encarta eine Bevorzugung amerikanischer Fragestellungen beobachtet wird.
- Weiterhin wird die mediale Aufbereitung der Artikel untersucht. Die Artikel der Encarta waren am umfassendsten mit Medienobjekten (Bilder, Audio, Video) ausgestattet, während die Wikipedia noch hinter dem Brockhaus zurückblieb. Erwartungsgemäß lag die Wikipedia hinsichtlich des internen Vernetzungsgrades vorne, gefolgt von Encarta und Brockhaus. Eine Absicherung der Information durch externe Literaturangaben war am häufigsten beim Brockhaus zu finden, gefolgt von Encarta und Wikipedia.

### 2.3 Die Studien von Stvilia, Twidale, Gasser und Smith [Stvilia et al. 2005, 2005a]

Die vorgenannten Studien gingen jeweils von einem impliziten Qualitätsmodell aus, welches die Grundlage für eine vergleichende Bewertung der *Wikipedia* bildete. Die Studien von Stvilia, Twidale, Gasser und Smith entwickeln zunächst ein Qualitätsmodell für die *Wikipedia* [Stvilia et al 2005]. Grundlage dieses Qualitätsmodells ist die in der (englischen) *Wikipedia* vorgenommene Hervorhebung von “Featured Articles”. Grundlage sind zunächst die Qualitätskriterien der *Wikipedia*: “Comprehensive, Accurate, Verifyable, Stable, Well-written, Uncontroversial, Compliance, Appropriate Images with acceptable copyright status, Appropriate Style and Focus.” Diesem Kriterienkatalog stellen sie einen von [Crawford 2001] vorgeschlagenen gegenüber, der sich auf traditionelle Print-Enzyklopädien bezieht: “Scope, Format,

Uniqueness, Authority, Accuracy, Currency, Accessibility.” Auch einen eigenen, umfangreicheren Kriterienkatalog, der hier nicht dargestellt werden soll, setzen sie zu den beiden anderen in Beziehung. Damit ist eine Grundlage für die Systematisierung von Qualitätsstudien sowie den Vergleich verschiedener Qualitätsmodelle gegeben. Im empirischen Teil der Studie stellen die Autoren sodann den Zusammenhang zwischen abstrakten Qualitätsmodellen und der Praxis der *Wikipedia* her. Dies erfolgt aufgrund einer inhaltlichen Analyse der den Lemmata zugeordneten Diskussionsseiten. Dazu wurden aus der *Wikipedia* und der Menge der Featured Articles jeweils 30 Artikel ausgewählt, die über einen hinlänglich umfangreichen Diskussionsteil verfügten. Die in der Diskussion auftretenden Argumente wurden kategorisiert und den Qualitätsdimensionen der *Wikipedia* zugeordnet. Somit wurde eine Einschätzung der Bedeutung der einzelnen Qualitätsaspekte für die Einordnung als Featured Article möglich. Zusätzlich wurde versucht, Textparameter zu isolieren, welche als Symptome für eine hohe Qualität der Artikel anzusehen sind (Umfang und Struktur der Diskussion, Umfang und Überarbeitungsfrequenz der Artikel, Anzahl der Autoren etc.). Dieser quantitative Ansatz wurde in einer weiteren Studie vertieft [Stvilia et al 2005a], so dass ein automatischer Qualitäts-Klassifikator entwickelt werden konnte.

#### 2.4 Weitere Studien mit Qualitätsbezug

Die Studie von Andrew Lih [Lih 2004] befasst sich mit der Verlässlichkeit der *Wikipedia* als journalistische Quelle. Er entwickelt eine einfache Metrik, welche die Qualität von Artikeln erfassen soll. Sie beruht auf der Anzahl von Änderungen, die ein Artikel durchlaufen hat und der Anzahl der beteiligten Autoren. Während eine zunehmende Zahl der Bearbeitungen für eine hohe inhaltliche Konsolidierung eines Beitrags spräche, erhöhe eine große Anzahl von Autoren die Wahrscheinlichkeit von Inkonsistenzen und Argumentationsbrüchen. Die Metrik wurde an einem Testkorpus kalibriert, um dann an Artikeln, die von der Presse zitiert wurden, getestet zu werden. Diese Studie initiierte den Einzug quantitativer Methoden in die Analyse der *Wikipedia*.

Emigh und Herring [Emigh und Herring 2004] vergleichen zwei elektronische Enzyklopädien (*Wikipedia* und *Everything2*) hinsichtlich des verwendeten Sprachregisters. Sie beobachten in *Everything2* eine vergleichsweise freie Sprache, während die *Wikipedia* zunehmend eine eher förmliche Sprache verwende. Wenngleich diese Studie keinen unmittelbaren Bezug zur Frage der Qualität der *Wikipedia* hat, ist doch von Interesse, dass der Redaktionsprozess eine Vereinheitlichung der Sprache

zur Folge hat. Dieser Effekt ist als relevant für Qualitätskriterien wie Lesbarkeit und sprachliche Konsistenz anzusehen.

## 2.5 Zusammenfassende Würdigung

Bisher sind noch vergleichsweise wenige Studien zur Qualität der *Wikipedia* erschienen. Diese sind zudem methodisch nicht unangreifbar. Auffällig ist die Diskrepanz zwischen den aus Studien resultierenden Qualitätsurteilen und den zum Teil witzigen Pressemeldungen über aktuelle, zum Teil krasse Fehlleistungen der *Wikipedia* [z. B. Rühle 2006]. Dies mag z. T. an methodischen Problemen der Studien liegen, hat seinen Hauptgrund aber vermutlich darin, dass diese Studien ihren Blick auf die mittlere Qualität einer Grundgesamtheit richtet, die Presse aber immer zu Recht auf besonders problematische Einzelfälle verweist. Darin liegt kein Widerspruch. Eine (traditionelle) Enzyklopädie muss sich der Anforderung stellen, dass auch der schlechteste Artikel gewisse Qualitätsschranken nicht unterschreitet. Insofern sind die oben genannten Studien – ebenso wie die noch folgenden – nur begrenzt geeignet, Werturteile über die *Wikipedia* oder andere Enzyklopädien zu begründen. Sie sind jedoch von Nutzen, die Effekte zu verstehen, die aus den neuen Redaktionsprozessen der *Wikipedia* resultieren. Besonders viel versprechend scheinen dabei solche Ansätze zu sein, die es erlauben, Qualitätsindikatoren automatisch für große Textmengen zu bestimmen. In diese Richtung verweisen die Arbeiten von Stvilia, Twidale, Gasser und Smith.

## 3 Die deutsche *Wikipedia* im Vergleich zu *Der Große Brockhaus*

Im Rahmen von Projektseminaren und Abschlussarbeiten wurden im Studiengang Informationswissenschaft der Universität Regensburg innerhalb des letzten Jahres einige Studien über die *Wikipedia* erarbeitet. Hier soll eine Studie vorgestellt werden, welche Qualitätsaspekte der deutschen *Wikipedia* mit solchen des *Brockhaus* vergleicht [Fuchs, Kattenbeck, Sax 2006]. Diese Studie ist sowohl hinsichtlich der Methodik als auch hinsichtlich der Fragestellungen sehr eng mit der zuvor vorgestellten Studie von Herrn Schlieker verwandt, geht aber in mancher Hinsicht über sie hinaus. Ausgewertet wurden die deutsche *Wikipedia* und *Der Große Brockhaus* (ebenfalls 20. Ausgabe). Die Stichprobe wurde nach dem gleichen Verfahren gezogen (random walk im Fall des *Brockhaus*, Zufallsartikel aus der *Wikipedia*), allerdings wurde ein Stichprobenumfang von 50 Lemmata für jede Enzyklopädie so ge-

wählt, dass sie auch in der jeweils anderen vertreten waren. Unterschiede im Ergebnis der Studien können demnach aus dem erweiterten Stichprobenumfang sowie dem erheblichen Anwachsen der *Wikipedia* (zwischen den Stichprobenziehungen liegt ein Jahr 8.2004 – 11.2005), deren Umfang sich in diesem Zeitraum mehr als verdoppelte [Wikipedia 2007c], resultieren.

### 3.1 Überprüfte Qualitätsaspekte

Zunächst soll kurz dargestellt werden, welche Qualitätsaspekte in den Vergleich einbezogen werden sollen. Es sei darauf hingewiesen, dass hier keine Auffassung hinsichtlich einer absoluten Interpretation dieser Qualitätsmerkmale vertreten wird.

Nicht überprüft wird in dieser Studie die sachliche Korrektheit der zu vergleichenden Enzyklopädieartikel. Zwar ist dies ein sehr wichtiges Qualitätskriterium, eine inhaltliche Überprüfung einer großen Anzahl zufällig ausgewählter Artikel wäre im Rahmen dieses Projekts jedoch nicht zu leisten gewesen. Im Rahmen eines gleichzeitig durchgeführten Schwesterprojekts wurde jedoch auch die sachliche Korrektheit von *Wikipedia*-Artikeln untersucht. Dabei waren nicht zufällig ausgewählte Artikel Gegenstand der Untersuchung, sondern es wurde eine Vollerhebung zu einem eingeschränkten Gegenstandsbereich durchgeführt, für den Beurteilungskompetenz vorhanden war (Shakespeares Werk) [Molz, Burghardt 2006].

Überprüft wurden in dieser Studie folgende Qualitätsaspekte:

- Zunächst wurde der gegenseitige Abdeckungsgrad der beiden Enzyklopädien festgestellt. Dies kann als eine Operationalisierung des Qualitätskriteriums „Vollständigkeit“ in Bezug auf die Gesamtzyklopädie angesehen werden.
- Um festzustellen, ob der Abdeckungsgrad thematische Abhängigkeiten aufweist, wurden die untersuchten Lemmata thematischen Kategorien zugeordnet, so dass festgestellt werden konnte, ob bestimmte Themen von einer der Enzyklopädien besser repräsentiert wurden.
- Die Vollständigkeit der einzelnen Artikel wurde heuristisch gemessen, indem einerseits die Artikellänge bestimmt wurde. Diese rein quantitative Angabe wurde ergänzt durch subjektive Einschätzungen.
- Die Überprüfbarkeit der in den Artikeln erhobenen Behauptungen wurde anhand der Anzahl der dem Artikel zugeordneten Quellenangaben und deren Korrektheit festgestellt.

- Als ein Indikator für die Korrektheit von Artikeln wurde die Sorgfalt angesehen, die in seine sprachliche Ausarbeitung investiert wurde. Hier wurden Orthographie, Interpunktion als leicht operationalisierbare Qualitätskriterien untersucht.

### 3.2 Thematische Abdeckung

Um 50 Lemmata des *Brockhaus* zu finden, die auch in der *Wikipedia* vorhanden waren, mussten 87 Stichworte gezogen werden. Umgekehrt waren 205 Ziehungen aus der *Wikipedia* erforderlich, um 50 Lemmata zu erhalten, die auch im *Brockhaus* repräsentiert waren. Zur Zeit der Stichprobenentnahme verfügte die *Wikipedia* bereits über 20% mehr Einträge als der *Brockhaus*. Es war also zu erwarten, dass die *Wikipedia* eine bessere Abdeckung erzielen würde. Dass die Abdeckungsquote für den *Brockhaus* noch ungünstiger ausfällt, dürfte auf die große Anzahl von Redirects zurückzuführen sein, die nicht in die Zählung der Lemmata der *Wikipedia* eingehen – ihre Zahl hat sich im Vergleichszeitraum verdreifacht [Wikipedia 2007c].

### 3.3 Umfang und Vollständigkeit der Artikel

Neben der rein thematischen Abdeckung ist auch von Interesse, in welcher Ausführlichkeit die Themen behandelt werden. Hier hat die Online-Enzyklopädie den Vorteil, dass im Gegensatz zur Buchenzyklopädie keine Einschränkungen hinsichtlich des Gesamtumfangs zu beachten sind. In der Stichprobe zeigt sich, dass 90 von 100 Artikeln des *Brockhaus* kürzer als 200 Worte sind, während 54 der *Wikipedia*-Artikel länger als 200 Worte sind. Der t-Test für verbundene Stichproben zeigt bei einem Niveau von 5%, dass die Längenunterschiede nicht zufällig sind.

Die Länge eines Artikels ist aber nur ein schwacher Indikator für seine inhaltliche Vollständigkeit. Schließlich kann ein vollständiger Artikel knapp formuliert, ein langer jedoch einfach hoch redundant sein. Als Annäherung an diese Problematik wurde für jedes Artikelpaar von 3 Personen – den drei studentischen Koautoren der Studie – ein Artikel als der vollständigere benannt. Es ist offensichtlich, dass drei Urteile für eine valide Studie zu wenig sind. Zudem waren die Juroren nicht unbefangen, da sie für jeden Artikel wussten, woher er stammte. Die Urteile wurden jedoch nur gewertet, wenn sie für einen Artikel einstimmig ausfielen. Unsicherheit besteht auch aufgrund der schwachen Definition des Begriffs „Vollständigkeit“. Letzterem Problem wurde begegnet, indem für die gefällten Urteile Cohens  $\kappa$  zur Ermittlung der Interrater-Reliabilität bestimmt wurde. Da der  $\kappa$ -Wert nur für ein Beurteilerpaar bestimmt werden kann, so wird er, nach [Roth 1984] für alle Beurteilerpaare ermittelt und der Median der Werte als durchschnittlicher Überein-



stimmungsgrad für alle Rater angenommen. Der resultierende Wert von  $\kappa=0,543$  kann in diesem Kontext als Zeichen für ausreichende Überstimmung der Beurteiler hinsichtlich des zu beurteilenden Kriteriums angenommen werden. Das Urteil fiel so aus, dass in 59 Fällen der *Wikipedia*-Artikel als vollständiger angesehen wurde, in 17 Fällen der *Brockhaus*-Artikel. Für die restlichen Fälle ergab sich kein einstimmiges Urteil. Dass hier ein eigenständiges Kriterium erfasst wurde, zeigt sich darin, dass nicht immer der längere Artikel als der vollständigere eingeschätzt wurde. Der Unterschied zwischen den beiden Enzyklopädien wurde mit Hilfe des Binomialverteilungstest bei einem Signifikanzniveau von 5% als nicht zufällig eingestuft. Damit kommt diesem Test, selbst bei Berücksichtigung der oben erwähnten methodischen Probleme, ein zumindest heuristischer Wert zu.

Mit der gleichen Vorgehensweise wurde auch noch versucht, zu einer Einschätzung der Verständlichkeit der Artikel zu kommen. Damit wäre ein weiteres, zweifelsohne wichtiges Qualitätskriterium erfasst worden. Es scheint aber, dass es über Verständlichkeit nur eine schwächere Intuition gibt als über Vollständigkeit. Jedenfalls konnte in diesem Fall keine hinlängliche Interrater-Reliabilität erzielt werden, so dass die diesbezüglichen Ergebnisse verworfen werden mussten.

### 3.4 Themenverteilung

Es wurde weiterhin versucht, festzustellen, ob die unterschiedliche Abdeckung der Gesamtmenge der Lemmata der Stichprobe auf inhaltliche Präferenzen der beiden Enzyklopädien zurückzuführen sein könnte. Dazu wurden zunächst die gemeinsamen Lemmata nach einem Ad-Hoc Kategoriensystem eingeteilt. Das Ergebnis ist in Abb. 1 zu sehen. Sodann wurden auch die Lemmata kategorisiert, die sich nur in der *Wikipedia* (Abb. 2) bzw. nur im *Brockhaus* (Abb. 3) fanden.

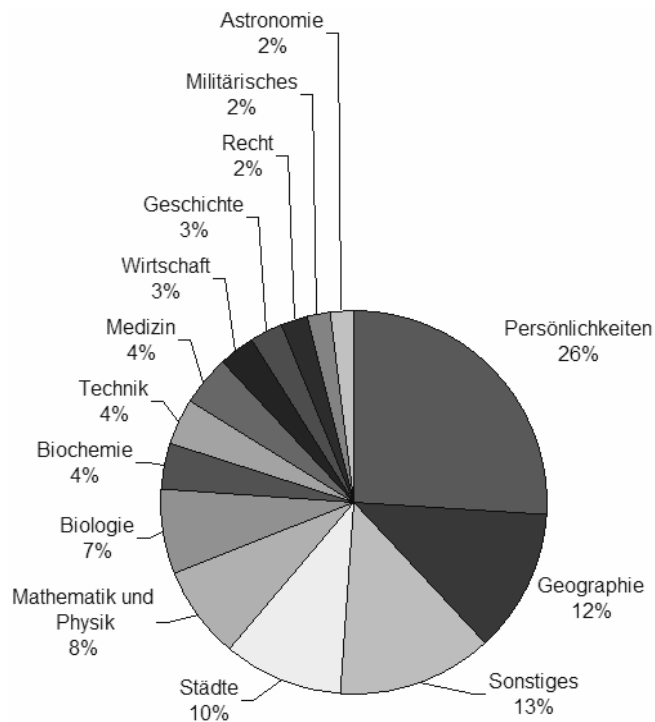


Abbildung 1: Kategorien der in beiden Enzyklopädien enthaltenen Lemmata

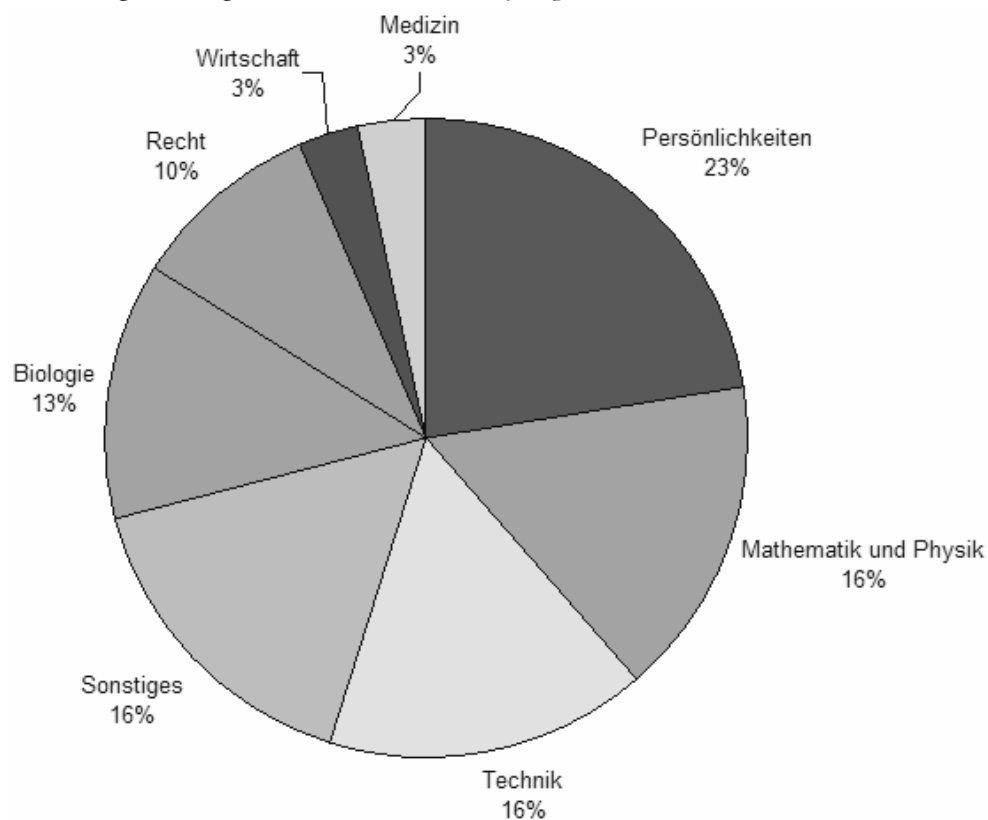


Abbildung 2: Kategorien der nur im Brockhaus enthaltenen Lemmata

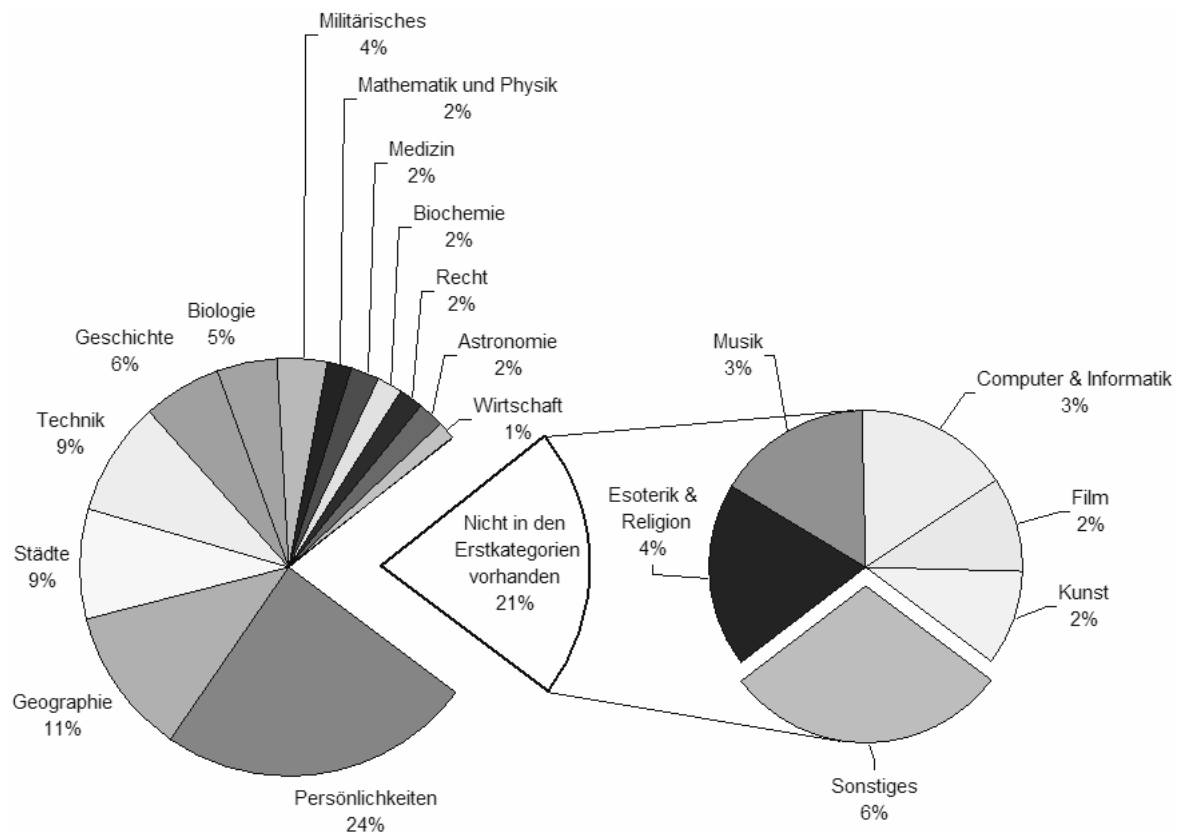


Abbildung 3: Kategorien der nur in der Wikipedia enthaltenen Lemmata

Während alle nicht in der *Wikipedia* auffindbaren Artikel unschwer in das ursprüngliche Kategoriensystem einzuordnen waren, mussten für die allein in der *Wikipedia* gefundenen Artikel einige neue Kategorien geschaffen werden. Bezüglich der kleineren Kategorien soll an dieser Stelle keine Aussage getroffen werden. Auffällig ist aber, dass die Kategorie Persönlichkeiten in allen drei Fällen etwa ein Viertel der Lemmata ausmacht. Biographische Information scheint beiden Enzyklopädien wichtig zu sein, Dissens besteht mitunter in der Auswahl der Personen. Der *Brockhaus* scheint eine etwas bessere Abdeckung im Bereich Naturwissenschaft und Technik zu erreichen, während in der *Wikipedia* Themen mit größerem Aktualitätsbezug Einzug gehalten haben (Computertechnik, Film, Musik) bzw. solche, die für traditionelle Enzyklopädien eher nicht ausgewählt würden (Esoterik).

### 3.5 Informationelle Absicherung durch Quellenangaben

Im Gegensatz zu den Beobachtungen von Schlieker sind in unserer Stichprobe die Artikel der *Wikipedia* besser durch externe Quellen abgesichert. 80 der *Brockhaus*-Artikel verfügten über keine Literaturangaben, aber nur 40 der *Wikipedia*. Zudem enthielten die *Wikipedia*-Artikel grundsätzlich mehr Quellenangaben. 344 Quellenangaben der *Wikipedia* standen 81 Angaben in den *Brockhaus*-Artikeln gegen-

über. Dieser Unterschied ist auf der Basis des t-Tests für verbundene Stichproben bei einem Signifikanzniveau von 5% als nicht zufällig anzusehen. Soweit wurden bei dem Test Online-Quellen und bibliographische Angaben gleich gesetzt, da sie beide dem gleichen Zweck dienen. Ein Unterschied wurde bei der Validierung der angegebenen Quellen deutlich. Während alle – rechnet man eine kleinere Unstimmigkeit nicht – Literaturangaben des *Brockhaus*, der auf die Nutzung von Online-Quellen verzichtet, sich als valide erwiesen, waren 10 der Online-Quellen der *Wikipedia* nicht auffindbar. Die bibliographischen Angaben waren in der *Wikipedia* ebenfalls valide.

Sehr auffällig war die hohe Anzahl von Artikeln, in denen Literaturverweise innerhalb des Artikeltexts nicht aufgelöst wurden (*Brockhaus*: 39, *Wikipedia*: 49). Dies stellt für beide Enzyklopädien einen erheblichen Mangel dar, wenngleich der T-Test für verbundene Stichproben zeigt, dass der Unterschied zwischen beiden Enzyklopädien signifikant ist.

### 3.6 Korrektheit in Orthographie und Interpunktion

Grundlage der Beurteilung sind die Regeln der neuen Rechtschreibung, auf die sich beide Enzyklopädien verpflichtet haben. Berücksichtigt werden selbst kleine Ungenauigkeiten, etwa falsches Setzen von Anführungszeichen. Es erwies sich, dass alle Artikel des *Brockhaus* fehlerfrei waren, während nur 31 *Wikipedia*-Artikel frei von Rechtschreibfehlern waren und nur 21 keine Interpunktionsfehler aufwiesen. Der Median des Auftretens von Orthographie- und Interpunktionsfehlern in der *Wikipedia* beträgt 2.

## 4 Zusammenfassung und Ausblick

Im Vergleich zu der Vorgängerstudie von Schlieker hat die *Wikipedia* innerhalb eines Jahres erhebliche Fortschritte in der thematischen Abdeckung und in der Qualität der informationellen Absicherung erzielt. Auch die größere Ausführlichkeit der Artikel ist eine Stärke der *Wikipedia*. Hier kann sie die Stärken des Online-Mediums ausnutzen. Die Schwächen in Interpunktion und Orthographie legen jedoch – selbst wenn es sich hier um sprachliche Sekundärtugenden handeln sollte – Schwächen im redaktionellen Prozess offen. Diese Fehler, wenn es auch nicht so viele waren, nähren den Verdacht, dass auch dort, wo komplexe Sachzusammenhänge redaktionell überprüft werden müssen, Mängel in der Sorgfalt bestehen könnten und sei es nur, weil es keine Zuständigkeiten gibt.

Deutlich wird in dieser Untersuchung auch, dass sie ein Zwischenstadium eines un-abgeschlossenen Prozesses erfasst. Zumindest die *Wikipedia* wird sich weiter verän- dern. Diese Veränderung wird dabei nicht nur quantitativer sondern auch qualitati- ver Natur sein. Dies zeigt sich schon in der hier dokumentierten neuen Rollenver- teilung hinsichtlich der informationellen Absicherung enzyklopädischer Informati- on. Auch für die traditionellen, d.h. nicht in offenen Redaktionsprozessen erstellten Enzyklopädien wird sich ein Druck zur Veränderung ergeben. Die weitere Entwick- lung wird auch beeinflusst von der künftigen Rolle von Alternativprojekten zur *Wi- kipedia*, wie etwa das von Larry Sanger propagierte Citizendium [Citizendium 2007].

## 5 Literaturangaben

- Citizendium (2007) The Citizendium Project. <http://www.citizendium.org/>, zitiert am 7.1.2007.
- Patrick Danowski und Jakob Voss (2005) Das Wissen der Welt – die Wikipedia. In: Open Source Jahrbuch. [http://www.opensourcejahrbuch.de/download/jb2005/chapter\\_o6/osjb2005-06-05-danowskivoss](http://www.opensourcejahrbuch.de/download/jb2005/chapter_o6/osjb2005-06-05-danowskivoss).
- Emigh, W., & Herring, S.C. (2005). In Collaborative authoring on the web: A genre analysis of online encyclopedias. Paper presented at the Proceedings of the Annual Hawaii International Conference on System Sciences. <http://ella.slis.indiana.edu/~herring/wiki.pdf>, zitiert am 6.1.2007.
- Karl-Peter Fuchs, Markus Kattenbeck, Christian Sax (2006) Brockhaus versus Wikipedia. Ausarbeitung zum Projektseminar Hypermedia im WS 2005/2006. Informationswissenschaft, Universität Regensburg.
- Jim Giles (2005) Internet encyclopaedias go head to head. *Nature* 438, S. 900-901, <http://www.nature.com/nature/journal/v438/n7070/full/438900a.html>, zitiert am 6.1.2007.
- Rainer Kuhlen (2005) Wikipedia – Offene Inhalte im kollaborativen Paradigma – eine Herausforderung auch für Fachinformation. *Forschung und Lehre*, Nr. 10, S. 546-548.
- Johannes Molz, Manuel Burghardt (2006) Shakespeares Werk in der Wikipedia. Eine bilinguale Untersuchung zu Umfang und Qualität von Wikipediaartikeln in einem abgegrenzten Themenbereich. Ausarbeitung zum Projektseminar Hypermedia im WS 2005/2006. Informationswissenschaft, Universität Regensburg.
- Jason Lanier (2006) Digitaler Maoismus. Kollektivismus im Internet, Weisheit der Massen, Fortschritt der Communities? Alles Trugschlüsse. <http://www.sueddeutsche.de/kultur/artikel/306/78228/article.html>, zitiert am 29.6.2006
- Andrew Lih (2004): *Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource*. In: Proceedings of the 5<sup>th</sup> International Symposium on Online Journalism.

- <http://jmsc.hku.hk/faculty/alih/publications/utaustin-2004-wikipedia-rc2.pdf>, zitiert am 6.1.2007.
- Erwin Roth (1984) Sozialwissenschaftliche Methoden. Lehr- und Handbuch für Forschung und Praxis. München et al. 1984.
- Alex Rühle (2006) Im Daunenfedergestöber. Wikipedia-Fälschungen. [sueddeutsche.de](http://www.sueddeutsche.de/kultur/artikel/63190541/), <http://www.sueddeutsche.de/kultur/artikel/63190541/>, zitiert am 6.1.2007.
- Christian Schlieker (2005) Explorative Untersuchung von Wissen in kollektiven Hypertexten, Diplomarbeit, Fachbereich 08, Soziologie, Universität Bremen.
- Besiki Stvilia, Michael B. Twidale, Les Gasser und Linda C. Smith (2005) Information Quality Discussions in Wikipedia. In: S. Hawamdeh (Ed.), *Knowledge Management: Nurturing Culture, Innovation, and Technology – Proceedings of the 2005 International Conference on Knowledge Management*. Charlotte, NC: World Scientific Publishing Company. S. 101-113. <http://mailer.fsu.edu/~bstvilia/papers/qualWiki.pdf>, zitiert am 6.1.2007
- Besiki Stvilia, Michael B. Twidale, Linda C. Smith, Less Gasser, (2005a). Assessing information quality of a community-based encyclopedia. In: *Proceedings of the International Conference on Information Quality – ICIQ 2005*. Cambridge, MA. S. 442-454. <http://mailer.fsu.edu/~bstvilia/papers/quantWiki.pdf>, zitiert am 6.1.2007
- Jakob Voss (2005) Informetrische Untersuchungen an der Online-Enzyklopädie Wikipedia, Magisterarbeit im Fach Bibliothekswissenschaft, Institut für Bibliotheks und Informationswissenschaft, Humboldt-Universität zu Berlin.
- Wikipedia (2007) Startseite der Wikipedia. <http://www.wikipedia.org/>, zitiert am 6.1.2007.
- Wikipedia (2007a) Forschungsportal der deutschen Wikipedia. <http://de.wikipedia.org/wiki/Wikipedia:Wikipedistik>, zitiert am 6.1.2007.
- Wikipedia (2007b) Wikipedia Research Bibliography. [http://meta.wikimedia.org/wiki/Wiki\\_Research\\_Bibliography](http://meta.wikimedia.org/wiki/Wiki_Research_Bibliography), zitiert am 6.1.2007.
- Wikipedia (2007c) Wikipedia Statistik Deutsch. <http://stats.wikimedia.org/DE/TablesWikipediaDE.htm>, zitiert am 7.1.2007

# Vom “Public Service” zum “Public Value”

## Öffentlich-rechtliche Programmarchive der Zukunft als digitale Wissensspeicher \*

*Dietmar Schiller<sup>1</sup>, Juliane Burghardt<sup>2</sup>,  
Nadine Fijalkowytsh<sup>3</sup>, Veit-E. Jauß<sup>4</sup>, Bodo Schindler<sup>5</sup>*

<sup>1</sup> Rundfunk Berlin-Brandenburg (rbb), Fernseharchiv, Leitung, *dietmar.schiller@rbb-online.de*

<sup>2</sup> Universität Potsdam, Institut für Linguistik, *burghardt@uni-potsdam.de*

<sup>3</sup> Freie Universität Berlin, Institut für Sprachwissenschaft; Universität Potsdam, Institut für Linguistik, *fijalkow@rz.uni-potsdam.de*

<sup>4</sup> rbb, Fernseharchiv, *veit-erdmann.jauss@rbb-online.de*

<sup>5</sup> rbb, Programmaustausch, *bodo.schindler@rbb-online.de*

### Zusammenfassung

Der Kurzbeitrag thematisiert den gegenwärtigen Wandel des Archiv-Verständnisses. Was bedeuten Programmvermögen als öffentliches Gut, Digitalisierung von Kulturgütern, offener Zugang zu digitalen Archiven, Verantwortung von Programmarchiven in der Wissensgesellschaft? Welche Rolle können im Zuge der fortschreitenden Digitalisierung künftig öffentlich-rechtliche Programmarchive in der Wissensgesellschaft übernehmen? Ausgehend von theoretischen Konzeptionen werden Überlegungen zu einem Programmarchiv der Zukunft angestellt, das sich dem Gemeinwohl verpflichtet fühlt und einen Mehrwert für alle darstellt.

### I Einführung

In den Archiven des öffentlich-rechtlichen Rundfunks lagert ein immenses Vermögen an Kulturgütern: schriftlich, auditiv, visuell. Seit Gründung der ARD vor fast sechs Jahrzehnten im Jahr 1950 ist in den Archiven der Landesrundfunkanstalten

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 91-96.

ein riesiger Bestand an Programm gewachsen: ein Spiegel der Gesellschaft, ein kulturelles Gedächtnis, ein Tagebuch der Alltagskultur. Dieses Programmvermögen zeichnet sich durch inhaltliche Vielfalt und Substanz aus – ein Speicher an unwiederbringlichen Dokumenten.

Aus dieser Perspektive stellen seine ständig wachsenden Dokumente einen zusätzlichen Wert für die Gesellschaft dar, der aber erst dann vollständig zur Geltung kommt, wenn er von allen Bürgerinnen und Bürgern auch genutzt werden kann: zeit- und ortsunabhängig. Deshalb stellt sich die Kernfrage, wie dieses gespeicherte Wissen einer offenen Gesellschaft zur Verfügung gestellt werden kann, und zwar jenseits der herkömmlichen Verbreitungswege?

Der öffentlich-rechtliche Rundfunk steht seit geraumer Zeit unter Druck. Im Wettbewerb mit den kommerziellen Mitstreitern geht es um eine Neu-Orientierung gegenüber den privaten Anbietern. Der durch Gebühren finanzierte öffentlich-rechtliche Rundfunk hat von Beginn an eine besondere Verantwortung für die Ausgestaltung einer demokratischen Gesellschaft übernommen, die sich in seinen Fernseh- und Radioprogrammen widerspiegelt. Gemäß des Funktionsauftrags lässt sich die historisch gewachsene und künftige Rolle der Programmarchive in gewinnorientiert (privat) und gemeinwohlorientiert (öffentlich-rechtlich) unterscheiden.

Die Struktur des öffentlich-rechtlichen Rundfunks in Deutschland hat sich stark an der BBC orientiert. Aus dem Grundverständnis der BBC als “public service“ lässt sich auch die grundsätzliche Aufgabenstellung ableiten, die die öffentlich-rechtlichen Programme in Deutschland zu erfüllen haben: Herstellung und Verbreitung gemeinwohlorientierter Programme für die gesamte Gesellschaft, die einen spezifischen Wert, einen Mehr-Wert (Public Value) für alle, darstellen.

## **2 Programmarchive der Zukunft: Zugang für alle und zu jeder Zeit!**

Ausgehend von dieser Prämisse lässt sich das im Wandel befindende Archiv-Verständnis mit folgenden Stichworten beschreiben:

- Programmvermögen als öffentliches Gut
- ein offener Zugang zu digitalen Archiven
- Verantwortung von Programmarchiven in der Wissensgesellschaft.

Diese Überlegungen basieren auf dem Konzept “Building public value. Reviewing the BBC for a digital world” (2004, S. 63), das ausgehend von einer Standortbe-



stimmung die künftige Rolle der BBC für die britische Gesellschaft umreißt. Dabei wird den Programmarchiven der BBC künftig folgende Aufgaben zugeschrieben:

The BBC Creative Archive will establish a pool of high-quality content which can be legally drawn on by collectors, enthusiasts, artists, musicians, students, teachers and many others, who can search and use this material non-commercially. And where exciting new works and products are made using this material, we will showcase them on BBC services.

Initially we will release factual material, beginning with extracts from natural history programmes. As demand grows, we are committed to extending the Creative Archive across all areas of our output.

We are developing this unique initiative in partnership with other major public and commercial audio-visual collections in the UK, including leading museums and libraries. Our ambition is to help establish a common resource which will extend the public's access while protecting the commercial rights of intellectual property owners.

Mit der Eröffnung des Fernseh Museums am Potsdamer Platz in Berlin im Frühjahr 2006 ist in Deutschland bereits ein erster Schritt getan. Zum Credo des Fernseh Museums gehört es, zusammen mit seinen Partnern, den öffentlich-rechtlichen und einigen kommerziellen Fernsehsendern, den historischen und kulturellen Wert des audiovisuellen Erbes in der öffentlichen Wahrnehmung zu verankern. Im Gegenzug heißt das für die Fernseharchive, sie dürfen auf lange Sicht nicht mehr nur dem Rundfunk selbst offen stehen, sondern allen, also der Öffentlichkeit; oder wie es der Intendant des NDR, Jobst Plog, im ARD-Jahrbuch 2005 (S. 22) aus Sicht der ARD formuliert: „Unsere Schlagkraft ist die Schlagkraft des Mehrwerts für alle. Diesen Mehrwert zu betonen und auszubauen, bedeutet, die Zukunft des öffentlich-rechtlichen Rundfunks zu sichern: wertvoll für jeden Einzelnen. Und: wertvoll für alle“.

Die Rolle der öffentlich-rechtlichen Programmarchive steht vor einem Paradigmenwechsel. Im Zuge der Digitalisierung audio-visueller Kulturgüter vollzieht sich auch eine Neuausrichtung des Aufgabenspektrums öffentlich-rechtlicher Programmarchive. Wie gestaltet sich der Wandel vom „analogen Archiv“ zum „digitalen Wissensspeicher“?

Fernseharchive sind ein Gedächtnis, in dem kollektiv bedeutsames Wissen der Gesellschaft gespeichert wird. Die Auswahl der Personen, Geschehnisse und Gegenstände, die Art ihrer Inszenierung und Präsentation beschreibt die Ansicht der Gesellschaft auf die Welt in einer Zeit. Fernsehproduktionen sind per se Kultur.

Digitalisierung und breitbandige Übertragungsmedien sind die Voraussetzungen für die Herstellung eines offenen Zuganges zu den Film- und Videomaterialien der Archive des öffentlich-rechtlichen Fernsehens. Während im Hörfunk die digitale,

filebasierte Speicherung der Produktionen schon seit längerem Alltag ist, so stellt sie sich beim Fernsehen als eine der großen, gegenwärtigen Aufgaben dar. Seit rund zwei Jahren werden die Produktionen zwar in digitaler Form als MPEG gespeichert, jedoch nach wie vor in konventioneller Form auf Magnetband gelagert.

Alle Produktionen sind im Format SDTV archiviert, welches nach und nach durch HDTV abgelöst wird. Der Bestand der Archive des öffentlich-rechtlichen Fernsehens besteht aus Materialien, deren Qualität professionelle Anforderungen erfüllt.

Die Verfügbarkeit der Archivmaterialien ist nach wie vor abhängig von konventionellen Transportdiensten und z. T. auch von der Anzahl vorhandener Kopien. In Zukunft können zwei Verfahren der Archivierung Möglichkeiten des Online-Zugriffs bieten:

- digitale, diskbasierte Speicherung auf Online-Server
- digitale, bandbasierte Speicherung in Bandrobotern

Während die diskbasierte Speicherung einen Online-Zugriff ermöglichen kann, wie er von der Nutzung des Internets her bekannt, wird der Zugriff bei der bandbasierten Speicherung eine Vorlaufzeit von einigen Minuten bis zum Beginn der Informationsübertragung aufweisen.

Was die Herstellung eines offenen Zuganges anbetrifft, so bietet sich die diskbasierte Speicherung für die Bereithaltung gering auflösender, wenig Speicherplatz benötigender Kopien an. Sie ermöglicht den jederzeitigen, schnellen Zugriff von allen. Diskbasierte Speicherung ermöglicht auch die Bereitstellung eines Programmbouquets, aus dem Kunden jeweils ihr individuelles Programm zusammenstellen können. Die bandbasierte Speicherung wird gehobenen Ansprüchen potentieller semi-/professioneller Nutzer gerecht, für die es in erster Linie um die Qualität und Verwertbarkeit der Materialien geht.

### **3 Programmarchive: Die Zukunft hat schon begonnen**

Wissen wird in einem Archiv gespeichert, damit bei Bedarf darauf zugegriffen werden kann. Durch Wissen können Daten und Informationen ziel- und zweckgerichtet erworben, verteilt und genutzt werden. Von zentraler Bedeutung für die Programmarchive öffentlich-rechtlicher Sender ist die Bereitstellung des gespeicherten Wissens für eine breite Öffentlichkeit.

Wissen bezeichnet auch zweckorientierte vernetzte Informationen, wobei computergestützte Arbeitstechniken Zugänglichkeit von Wissen überhaupt erst möglich

machen und zur Multimedialisierung des Datenaustauschs beitragen. Dabei ist zu beachten, dass das Thema Vernetzung vom technischen Medium Internet kaum zu trennen ist. Die für das Fernsehen oder Radio produzierten Programmbeiträge sollten trotz ihrer hohen Produktions- und Vervielfältigungskosten auf Senderseite einem großen Personenkreis zugänglich gemacht werden: einfach in der Bedienung, kostenneutral und barrierefrei.

Ein Beispiel für Mehrwert und eine nutzerbezogene Aufbereitung von Informationen bieten die elektronischen Publikationskataloge an deutschen Universitäten, wobei so genannte „e-journals“ zu Hypertexten durch Links vernetzt und die Datenbanken als Medienverbund organisiert ist.

Ein offener Zugang zu den öffentlich-rechtlichen Programmarchiven auf Basis eines digitalen, dezentralen Broadcast-Netzes würde die Bereitstellung audiovisueller Informationen in Preview-Qualität ermöglichen. Jede partizipierende Anstalt hätte somit die Möglichkeit, ihr archivierte und dokumentierte Wissen im Sinne des so genannten „Open-source“-Konzepts der Gesellschaft zur Verfügung zu stellen (vgl. Creative Access 2006). Vorstellbar wäre eine Art Netzwerk, um schnell und effektiv nach gewünschten Programmbeiträgen zu suchen und diese zu verwenden.

Der praktische Nutzen wäre immens: Das audio-visuelle Wissen der öffentlich-rechtlichen Programmarchive wäre somit für alle zu jeder Zeit abrufbar. Bisher nur selten verwendetes Material könnte für die Wissensgesellschaft nachhaltig genutzt werden. Ob Universitäten, Schulen, Bildungseinrichtungen, ob zur individuellen Weiterbildung, beruflichen Umorientierung, Horizonterweiterung – die in den Programmarchiven lagernden audio-visuellen Bestände rücken somit ins Zentrum der Wissensgesellschaft. Sie sind dem Gemeinwohl verpflichtet und stellen einen Mehrwert für alle da.

## **4 Offene Fragen**

Eine diesbezügliche Öffnung der Archive zieht eine Reihe von Fragen nach sich. Viele Aspekte zur Ausgestaltung dieses Szenarios sind noch nicht diskutiert. Dazu zählen die rechtliche und ökonomische Dimension ebenso wie ein öffentliches Marketing. Welchen Einfluss wird die Öffnung der Archive auf die im Entstehen begriffene Wissensgesellschaft haben? Welche Zielgruppen werden davon Gebrauch machen? Diese Fragen sollten – auf Basis der britischen Erfahrung – diskutiert werden.

## **5      Literatur**

- British Broadcasting Cooperation (BBC) (2004). Building Public Value. Renewing the BBC for a digital world. 2004. [http://www.bbc.co.uk/pressoffice/pressreleases/stories/2004/06\\_june/29/bpv.shtml](http://www.bbc.co.uk/pressoffice/pressreleases/stories/2004/06_june/29/bpv.shtml) [Zugriff März 2007]
- Demokratiezentrum Wien: Creative Access. Digital Archives Between Open Knowledge Society and Commodification in Vienna. 2006. [http://www.demokratiezentrum.org/de/startseite/projekte/abgeschlossene\\_projekte/creative\\_access.html](http://www.demokratiezentrum.org/de/startseite/projekte/abgeschlossene_projekte/creative_access.html) [Zugriff März 2007]
- Plog, Jobst (2005). "Mehr Wert für alle. Rundfunk ist ein öffentliches Gut – Warum die Zukunft der ARD auf diesem Gedanken ruht". In: Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland (ARD) (Hrsg.). ARD Jahrbuch 05. Hamburg: Verlag Hans-Bredow-Institut / Baden-Baden: Nomos, 15-22 [online unter: <http://www.ard.de/-/id=224808/property=download/14fbhc8/index.pdf>, Zugriff März 2007].

# OPEN MANTRA for Open Access to Information\*

*A.R.D. Prasad, Devika P. Madalli*

Documentation Research and Training Centre,  
Indian Statistical Institute  
Bangalore 560059, Karnataka, INDIA  
*{ard, devika}@drtc.isibang.ac.in*

## Abstract

The importance of Open Access (OA) to information cannot be over emphasized. If ‘Knowledge for all’ has to be a reality then ‘Information for all’ must be ensured. Though the concept of OA is simple and straight forward enough, it encompasses many challenges such as copyright, government policies, awareness, technology among several other issues. This paper discusses a few of these issues and possible solutions towards OA. The paper proposes the OPEN MANTRA<sup>1</sup> – Open SOURCE tools based upon Open STANDARDS to achieve OPEN ACCESS to information.

## I Introduction

In the scholarly world, access to information has a significant role in the production of information. Not all libraries can afford to acquire all resources needed by their patrons given the rising costs of publications. Further, unless we get access to the research output of other researchers, research productivity gets hampered; additionally, the chances of duplication of efforts would be high. To ensure that information reaches its patrons without barriers of cost or copyright, Open Access (OA) to information is essential. Internet and e-publishing have added a new dimension to the user’s expectation towards the open access to information. Publication of scholarly literature on Internet is simpler than traditional modes of publishing. Ease

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 97-108.

<sup>1</sup> Mantra in Sanskrit means chant or spell.

of use of HTML and availability of open source tools for e-publishing, helped in the emergence of more than 2500 open access journals and magazines.

As the OA movement is gaining momentum, more tools and technologies are coming up which make practical OA system possible. Open source digital library software (e. g. DSpace, EPrints) are the key enabling software to set up OA repositories. Software like Open Journal System (OJS) provide open access platform to publish e-journals incorporating many of the methodologies of print journals with regard to peer reviewing. Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH) is an excellent example of a facilitating technology enabling accessing of information across open access digital repositories and OAI – PMH compliant journals.

However, there are several serious challenges in the way of making the scholarly information open to all. In this paper we discuss some of the major challenges in achieving open access to information.

### **1.1 Peer-reviewing and its alternatives**

In general, the journal ranking is established on the basis of the peer review process. Though some of the OA journals are peer-reviewed, there is a common opinion that OA journals are NOT peer-reviewed and hence they are of low quality compared to commercially published journals. The underlying strength of peer review is “...the concerted effort by large numbers of researchers and scholars who work to assure that valid and valuable works are published and conversely to assure that invalid or non-valuable works are not published [...]”. (Weller, 2001).

Though many of us agree with the views of Weller about peer reviewing, it should be pointed out that peer reviewing is not without its pitfalls as expressed in literature and discussions.

- Some editors reject without sending the paper to referees
- Sometimes editors may deliberately choose a harsh referee for a paper the editor wishes to see rejected
- Bias against authors depending on nationality, gender, institution etc.
- Author and referee belonging to opposing school of thought
- Reviewers can reject a good paper and write a paper using the ideas
- Delaying publication of potentially competing research
- Reviewers may be good in the subject and may not have skills in reviewing
- The limitations of the subject knowledge of the reviewer himself may make him sceptical of the publication under review.

It should also be pointed that there are alternatives to peer-reviewing which cover some of the pitfalls of peer reviewing. After all, it is finally the reader who decides what to accept and what not to accept when he reads a publication. The following are some of the alternatives to peer reviewing though they too are not without pitfalls. However, the institution which intends to publish may look at the appropriateness of these alternatives.

- Certification-based (reviewers are trained)
- Open Peer Review (reviewers sign)
- Commentary-based (readers can comment before or after actual publication)
- Collaboratively-filtered (guiding readers on what to read)
- Institution-based (institutional repositories)
- Citation-based
- NO Peer Review

## 1.2 Mandate

Though the philosophy to OA appeals to most academicians and researchers who are most content generators, OA repositories and journals are not populated easily. Many a project in OA suffers from paucity of volunteered content. The authors have been active in campaigning OA in Indian scientific and academic community. It is often found that there is not enough motivation for authors that want to go Open Access. Performance assessment and career development still largely depends on how many published articles a person has in ‘ranked’ journals. So the need of the hour for sustained OA systems is MANDATE. Each institute should mandate that its research output should be deposited in their own Open Access Repositories.

The argument of OA activists world over has been that “output of publicly funded research must be made freely, publicly accessible”. This is in general the message brought by the Berlin declaration (Berlin Declaration, 2003), and supported by initiatives like the Budapest Open Initiative (Budapest Open Access Initiatives, 2001). The bills such as the *American Center for CURES Act* of 2005 and Federal Research Public Access Act of 2006 (FRPAA) (Peek, 2006) (English & Suber, 2006), tabled in US senate in support of making free access to publically funded research results, if passed, will greatly pave the way of OA to information.

## 2 Open Source tools and technology for OA

Open source movement has indeed contributed valuable tools for different applications and fortunately a good number of them can be used to offer products as OA resource repositories and archives. Digital Repositories are popular means of making available, institutional, domain based and community resources in an organized manner as openly accessible. There are quite a few very good tools that help build and manage digital repositories. We can broadly categorize the software in the following groups: (i) software for OA repositories; (ii) software for publishing and managing the journals; (iii) Metadata Harvesting software and (iv) Software for Digital Preservation. In the following paragraphs, we have discussed a few of the above mentioned software.

### 2.1 Digital Repositories

Open source software have facilitated the OA repositories in many ways. Popular open source digital repository tools include DSpace (<http://www.dspace.org>), EPrints (<http://www.eprints.org>), Fedora, GSDL ([www.greenstone.org](http://www.greenstone.org)), Fedora, CDSware etc. However, it is important that digital library managers make informed decisions of what tool would be suitable for their collections, patrons and services they plan to offer. Some of these softwares are broadly evaluated by us and we have chosen DSpace for our production system, Librarian's Digital Library (LDL – <https://drtc.isibang.ac.in>).

LDL is India's first DSpace based repository and only second in world in the domain of Library and Information Science (LIS). It has community collection building modus with authors from about 14 countries and has over 300 members. The activities are supported by Documentation Research and Training Centre (DRTC), Indian Statistical Institute (ISI) and through an associated discussion group, Digital Library Research Group, (DLRG). Initially, we have uploaded many articles published in the DRTC's annual and refresher seminar volumes of the recent past with born digital versions of the articles. Digitization of the older seminar volumes has been undertaken and many more publications will be uploaded soon. Though, we have invited many authors to upload their publications, if copyleft, we only had a lukewarm response. We believe, the reasons are varied, viz. fear of copyright laws mostly out of unawareness of copyright issues and also lack of awareness of open access movement, misplaced feeling that it requires knowledge of web to upload a document etc.



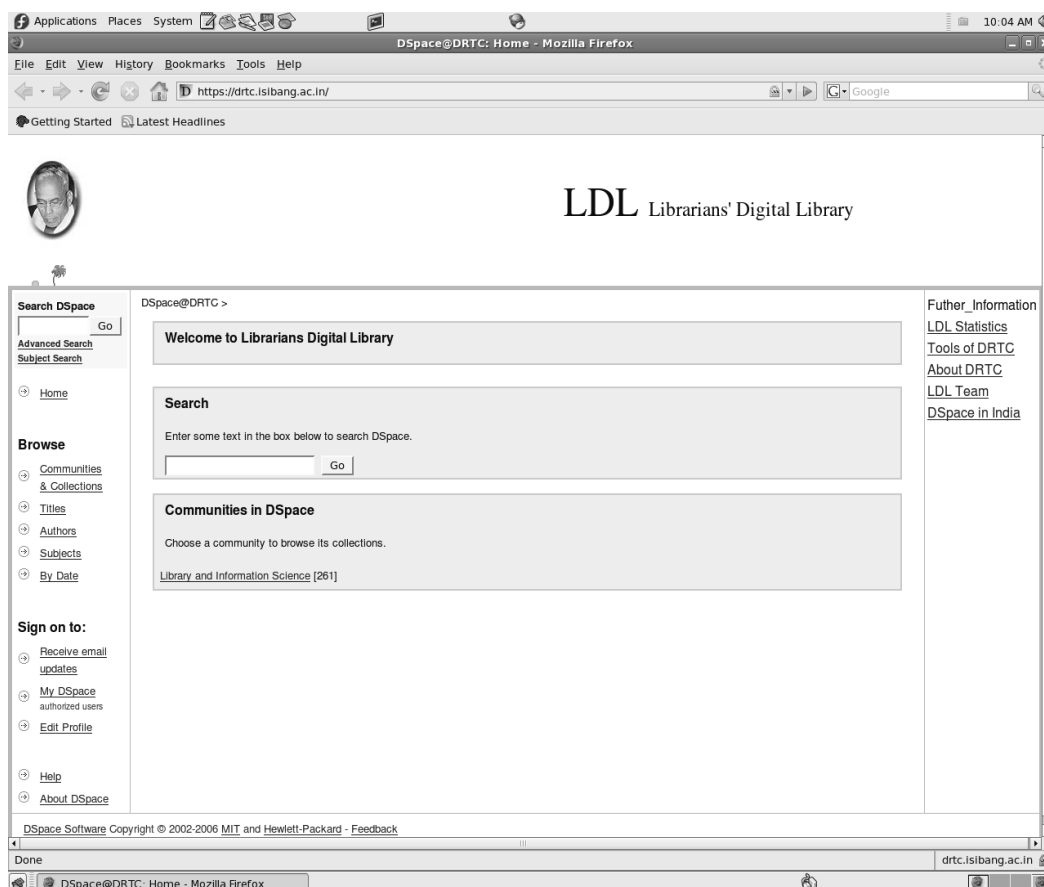


Figure 1: DRTC/ISI LDL Repository Service in LIS

## 2.2 Directory Services

**OpenDOAR (The Directory of Open Access Repositories):** OpenDOAR (<http://opendoar.org>) categorizes and lists wide variety of Open Access research archives. The project is a joint collaboration between the University of Nottingham in the UK and the University of Lund in Sweden. Both institutions are active in supporting Open Access to Information. We estimate about 1900 repositories and openDOAR covers about 836. OpenDOAR has an excellent search facility.

**Directory of Open Access Journals (DOAJ):** There are at present more than 2526 journals in the directory. Currently 753 journals are searchable at article level. As of today 125067 articles are included in the DOAJ (<http://www.doaj.org>) service. All the OA journals are categorized under broad disciplines, in addition to browsing and searching for a specific journal.

### Harvester Services

OAI-PMH ensures interoperability among complaint Digital repositories. Using OAI Harvester software many information services have come into existence. One of the best known such services is OAIster (<http://oaister.umdl.umich.edu/o/oaister/>).

Arxiv is yet another service popular in the field of physics, mathematics, computer science etc. At Documentation Research and Training Centre (DRTC), Indian Statistical Institute (ISI), we are providing a harvester server called Search Digital Libraries (SDL – <http://drtc.isibang.ac.in/sdl/>). SDL harvests metadata from various digital repositories in the field of Library and Information Science and presently provides access to more than 20,000 full text documents from 17 repositories.

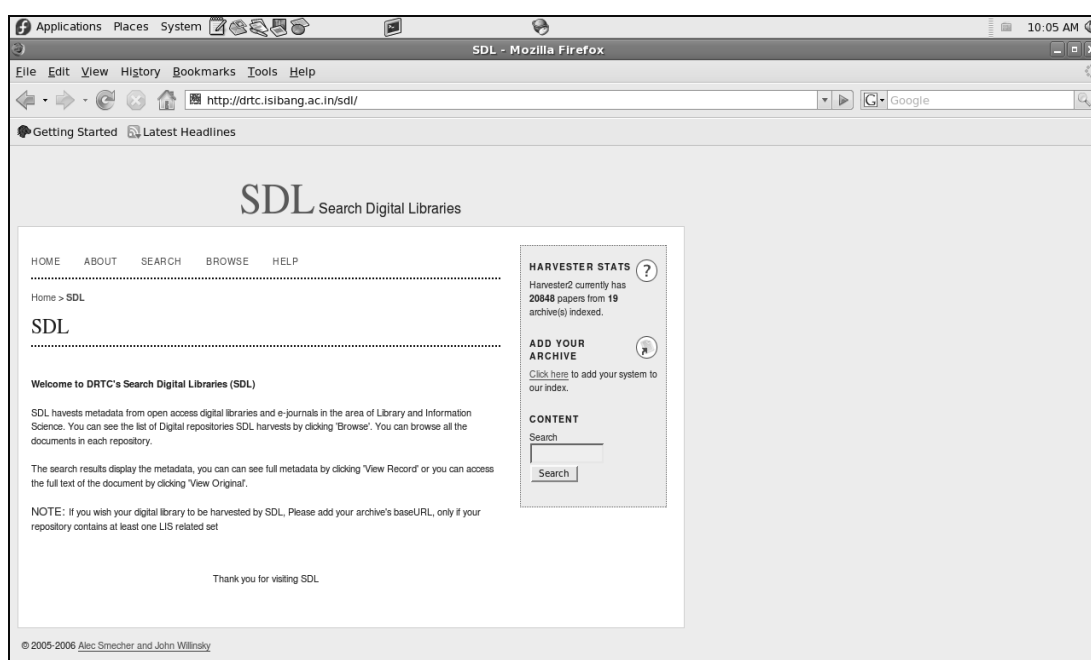


Figure 2: DRTC/ISI SDL Service in LIS

Though we know that there are more than 17 repositories having OA content in Library and Information Science (LIS), we have serious problems in harvesting those repositories. The methodology, we have adopted to identify repositories dealing with LIS content, is to go through the directory services like openDOAR, website of OpenArchives (<http://www.openarchives.org>). In addition, the software websites provide list of repositories that are using their software. For example, EPrints and DSpace sites do have list of organizations and the URLs of their repositories. However, it should be noted that these directories are not exhaustive as many people do not post information about their repositories in these directories. That means there is an urgent need to develop some mechanism or at least a convention to identify digital repositories and their URLs. Even when we identify a repository, identifying the baseURL in order to harvest metadata is yet another challenge. If the repository is using popular software like DSpace or EPrints one can guess the baseURL, even if a repository and its baseURL is not available in any of the directory services. However, in case of a significant number of repositories guessing the baseURL turns out to be a tedious and sometimes impossible effort. After all, one

can expose the metadata of repository using OAI-PMH in various ways using varied URL syntax using a PERL and PHP etc. programs.

In some cases, it is easier to identify the repositories dealing with LIS, especially if they deal exclusively with LIS, as in the case of DLIST, E-LIS, LDL, Australian Library and Information Association. However, it should be noted that many repositories, especially in the case of repositories host by universities, they may host many other subjects in addition to library and information science. The OAI-PMH protocol allows selective harvesting by 'set' or by 'date'. In case of some repositories, the LIS related content is normally available under a set. Using selective harvesting facility of OAI-PMH, one can get the list of sets in particular repository and harvest only the sets that deal with the subject or discipline of interest. The difficulty with selective harvesting is that many repositories might not have organized the content by subject. For example, if they create a set of 'Theses' and upload all theses from various subject under the set 'theses', the selective harvesting approach will not be of any use, as we can not filter out theses of other disciplines.

As there is no provision for selective harvesting by subject using OAI-PMH verbs, one can filter the metadata records after harvesting, that is at database level, as the harvested metadata is normally stored in a database. For example, PKP harvester has provision to store metadata records either in MySQL or PostgreSQL. The problem with this approach is that many people enter keywords following either a thesaurus or the keywords given in the publication or adhoc keywords. Even in case of controlled vocabulary they may follow different thesauri to enter the keywords and this does not ensure uniformity in the choice of keywords.

One way, we believe, to overcome this problem is to add a new element to Dublin Core metadata schema viz. dc:discipline and to mandate that one should choose the discipline from Library of Congress subject heading or Sears list either up to 1<sup>st</sup> order or 2<sup>nd</sup> or 3<sup>rd</sup> order array in the hierarchy of subjects, along with the broader terms. In addition, OAI-PMH protocol should have a selective harvesting by 'discipline'. Then it would be easier to build harvester for various disciplines. With the present OAI-PMH and Dublin Core versions, it is difficult to build satisfactory discipline based harvester services and DRTC's SDL suffers from these lacunae, as do any other domain-based harvester service.

### **2.3 Journal Publishing**

Open Journal System (OJS), HyperJournal, DPubs are some examples of OSS tools for publishing and managing the OA scholarly journals and e-monographs. Over

800 OA journal titles (as of June 2006) in ten languages are already using OJS as OA (OJS-Journals). Many of the journal publishing software support OAI-PMH protocol. Though DOAJ lists about 71 journals in LIS, we have identified more than 80 electronic journals in the field of library and information science. Unfortunately, DOAJ has not tackled issue of baseURLs and we could not get the baseURL of these journals to harvest the contents of the journals in our SDL service. When we wrote to these journals, some did not respond and some of them wrote back their site is not OAI-PMH compliant.

The problem, we believe, is that many journals were initiated much before the OAI protocol came into existence. Though, there are tools available in open access like OCLC's OAIcat to expose the metadata and turn the site into a data provider, still the task of entering metadata for every article published by these journals is a problem to reckon with. In addition to OAI protocol, the RSS feed is yet another way of providing the contents list of journals which a significant number of journals are providing. However, harvesters using OAI protocol do not have mechanism to capture metadata from these journals. An interesting problem worth investigating is possible to have the harvester software to embed RSS facility into their software.

DOAJ site has been made OAI compliant. Of the 2549 journals listed in DOAJ, 757 journals' articles can be searched at article level and total of 125716 articles metadata is available. Unfortunately, the problem is that when we use OAI verb to harvest metadata of these, it gets metadata of all the articles and there is no way to do selective harvesting by journal title or ISSN. We hope in future, these issues will be address for better retrieval of information available in open access journals.

## **2.4 Open Standards**

The commercial competitive world of computer software products and tools often push the standard which in turn makes certain tool inevitable. Ideally, OA models should be able to develop, grow, change and sustain over a period of time. The direct implication is that the solutions over a period of time should be compatible through the versions of software while being cost effective. Also the changes and migration should not entail buying new software or entering into more agreements and licensing for their use. OA systems are often made available through collaborative development online across communities and probably to as many patrons as are interested in the content. Given this situation, Open Source Software seem the ideal which while being free for download and use, provide autonomy in operations of OA collaborative projects. But there is need for caution while choosing the OSS. It is not only enough if it is FREE but it is equally important that it is NOT based on

or using at any stage proprietary standards. To implement OA successfully it is important the OSS be only based on OPEN STANDARDS.

There is no universally accepted definition available for open standards. Many people have different views about it. In general, open standards are publicly available, independent of any single institution or manufacturer. The key characteristics of open standards as identified by Coyle (Coyle, 2002) are: 1) that anyone can use the standards to develop software, 2) anyone can acquire the standards for free or without a significant cost, and 3) the standard has been developed in a way in which anyone can participate. The Danish Government defines (eGovernment News, 2004, 28 June 04) open standards as, it should,

- Be accessible to everyone free of charge: no discrimination between users, and no payment or other considerations should be required as a condition to use the standard,
- Remain accessible to everyone free of charge: owners should renounce their options, if any, to limit access to the standard at a later date.
- Be documented in all its details: all aspects of the standard should be transparent and documented, and both access to and use of the documentation should be free.

It is important for libraries and other cultural institutions to ensure long-term access to digital information. The rapid growth in digital technologies has led to new and improved applications for digital preservation. However at the same time it has also led to some problems as well. Two of these problems are obsolescence and dependency issues. The obsolescence problem is caused by the advances in hardware and software making many computers obsolete within three to five years (Vilbrandt, et al. 2004). Dependency problems can arise if tools that are needed to communicate between systems or read file formats become unavailable. In order to account for obsolescence and dependency problems organizations must be able to migrate data into new systems. Data migration, however, cannot occur without access to data file formats.

Some important characteristics of open standard can be identified as,

- It should not be encumbered by a patent
- It should not require proprietary software
- Can be utilized by anyone without cost
- It should be available for all to read and implement
- Implementations of Open Standards may be extended, or offered in subset form

- It should create a fair, competitive market for implementations of the standard. Should not lock the customer in to a particular vendor or group.

The categories of standards that directly having a bearing on OA broadly are:

1. File formats
2. Digital Preservation Standards
3. Metadata and data standards
4. Technical and Information Architecture standards
5. Interoperability standards
6. Encoding Standards
7. Network Protocols and standards

Examples of some open standards are: SQL (a specification approved by ANSI and ISO), HTML/XHTML (specifications of the W<sub>3</sub>C for structured hyperlinked document formatting), OpenDocument (a specification by OASIS for office documents, approved by ISO as ISO/IEC 26300), etc.

### 3.5 Changing Copyright Scenario

SHERPA's RoMEO project (<http://www.sherpa.ac.uk/romeo.php>) categorizes publishers into different color categories according to their policies regarding copyright and self-archiving. Green category lists those publishers which allow authors to submit the preprints and post prints to self archives. Elsevier, Emerald, Haworth Press, IEEE, John Wiley, Sage, and Springer are among the major commercial publishers who have accepted the self-archiving policies applicable of RoMEO Green category with certain conditions, such as (SHERPA – RoMEO):

- Author can submit to personal or author's institutional server/ repository
- Published source must be acknowledged
- Must link to journal home page
- Publishers version/PDF cannot be used
- Articles in some journals can be made Open Access on payment of additional charge

## 4 Conclusion

It is the reviewers, who are domain experts and the reviewing process that makes a journal a highly ranked one and it will take a few years to consider a journal to be a good quality one. Considering the fact that commercial journals have been in exis-

tence more than 100 years and OA journals are hardly a decade old, it is too optimistic to expect OA will pervade soon. But it is difficult to deny the fact that the results of OA movement – OA journals, OA repositories is very impressive and significant.

The key to quick success of OA is highly dependent on the mandates from the authorized bodies. This will encourage the scholars to publish their publications in open access journals and to deposit it to institutional or self-archiving repositories. It is highly inadvisable to mandate direct deposit in a Central Repository (CR) -- whether discipline-based, funder-based, multidisciplinary or national. The right way to get OA content into CRs is to harvest it from the IRs (via the OAI protocol) (Harnad, 2006).

Unfortunately, many authors are apprehensive of the copyright issues. As many commercial publishers are slowly yielding to the pressure of OA movement, authors should be aware of the freedom of publishing their research work. There is nothing wrong with copyright as long as rightly practiced. It is the 'transfer of rights' that authors should concern themselves with. In any case, the author should only grant a non-exclusive right to publish and disseminate their articles. During the period of euphoria of the emergence of democracies, many founding fathers of democracy believed that free (universal) education will help in building nations. Similarly, universal access to information will help in building a better world bridging the gap between so called developed and developing nations.

## References

- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.  
Conference on Open Access to Knowledge in the Sciences and Humanities held from 20 – 22 October, 2003, Berlin. <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>
- Budapest Open Access Initiative. Meeting organised by the Open Society Institute (OSI) held at Budapest on December 1-2, 2001. <http://www.soros.org/openaccess/>
- Coyle, K. 2002. Open source, open standards. *Information Technology and Libraries* 21(1): 33-36.
- eGovernment News – 28 June 2004 – Denmark/EU & Europe-wide – Interoperability. <http://europa.eu.int/idabc/en/document/3132/333>
- Harnad, Stevan (2006). Optimizing OA Self-Archiving Mandates: What? Where? When? Why? How? <http://eprints.ecs.soton.ac.uk/13098/01/arch.html>.
- LOCKSS: How it works. [http://www.lockss.org/lockss/How\\_It\\_Works](http://www.lockss.org/lockss/How_It_Works)
- Open Access News: Who, what & why. BOAI. <http://www.biomedcentral.com/openaccess/www/?issue=4>

Open Access Now: What, Who and Why?. SPARC.

<http://www.biomedcentral.com/openaccess/www/?issue=3>

Ray English and Peter Suber. (June 2006). Public access to federally funded research: The Cornyn-Lieberman and CURES bills. C&RL News, June 2006. Vol. 67, No.

<http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2006/june06/fedfundedresearch.htm>

Rhyno, Art. Using open source systems for digital libraries. Westport: Libraries Unlimited, 2004. pp26.

Robin Peek (January 2007). The Federal Research Public Access Act of 2006. Information Today, January 30, 2007. <http://www.infotoday.com/newsbreaks/nb060508-2.shtml>

Vilbrandt, T., et al. 2004. Cultural heritage preservation using constructive shape modeling. Computer Graphics Forum 23(1): 25-41.

Weller, Anne C. (2001). Editorial Peer Review: Its Strengths and Weaknesses. Medford, NJ: Information Today.



# Aufbau und Betrieb digitaler Repositorien in Indien

## Eine empirische Studie am Beispiel Open Source basierter Repositorien\*

*Achim Osswald*

Cologne University of Applied Sciences  
Institut für Informationswissenschaft  
Claudiusstrasse 1, 50678 Köln  
*achim.osswald@fh-koeln.de*

### **Zusammenfassung**

Institutionelle oder themenorientierte digitale Repositorien bilden zunehmend einen wichtigen Beitrag zur Verbreitung wissenschaftlicher Erkenntnisse. Sie werden weltweit mehrheitlich mit Open Source Software betrieben. Auf der Grundlage einer empirischen Studie aus dem Jahr 2006 gibt der Beitrag einen Einblick in die Rahmenbedingungen und den aktuellen Stand beim Aufbau und Betrieb solcher Repositorien in Indien.

### **Abstract**

Institutional and subject repositories are getting growing importance for science related communication. Worldwide Open Source Software is used for setting them up. Based on an empirical study in 2006 the paper gives insights in the circumstances of setting up and maintaining such repositories in India.

## **I Einführung**

Das Konzept Digitaler Bibliotheken ermöglicht die Bereitstellung von digitalen Ressourcen als thematische, institutions- bzw. organisations- oder dokumenttypbezogene Sammlungen digitaler Objekte. Neben den international renommierten, themenorientierten Repositorien (*subject repositories*) wie dem berühmten Ar-

---

\* Veröffentlicht in: Osswald, Achim; Stempfhuber, Maximilian; Wolff, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 109-127.

Xiv.org<sup>1</sup>, wurden im deutschsprachigen Raum schon relativ frühzeitig (ca. ab 1998) institutionelle Repositorien (IR) zumeist als sog. Hochschulschriftenserver eingerichtet. Für deren Betrieb wurden mittels öffentlicher Fördergelder deutschsprachige Open Source Programme entwickelt.<sup>2</sup> Ende 2006 haben von diesen Entwicklungen in Deutschland OPUS<sup>3</sup> mit derzeit 55 sowie MyCoRe<sup>4</sup> mit derzeit 14 Anwendungen die größte Bedeutung. Open Source Produkte des internationalen Angebots haben in Deutschland bislang nur wenig Bedeutung.

DINI, die Deutsche Initiative für Netzwerkinformation e. V. bemüht sich seit Jahren, um das funktionale Niveau der bestehenden Publikationsserver. Hierzu wurde im Jahre 2003 und in 2. Auflage im September 2006 das „DINI-Zertifikat Dokumenten- und Publikationsservice 2007“ [DINI 2007] erstellt und propagiert. Die dort aufgestellten qualitativen Beurteilungskriterien sind:

- Sichtbarkeit des Gesamtangebotes
- Leitlinien (Policy)
- Autorenbetreuung
- Rechtliche Aspekte
- Integrität
- Erschließung
- Zugriffsstatistik
- Langzeitverfügbarkeit

Im Jahr 2006 wurde das Zertifikat auch in Indien als qualitative Orientierung in die Diskussion gebracht [Oßwald 2006]. Auch wenn es nach Ansicht indischer Exper-

---

<sup>1</sup> ArXiv.org (genannt "The Archive") ist ein Archiv für Preprints (auch e-prints oder eprints genannt) aus den Bereichen Physik, Mathematik, Informatik und Biologie ursprünglich am Los Alamos National Laboratory (LANL) gehostet, jetzt wird es an der Cornell University betrieben und weltweit gespiegelt.“ <http://de.wikipedia.org/wiki/ArXiv> (Stand dieser und der anderen Webadressen vom 16.1.2007)

<sup>2</sup> Zur terminologischen Klärung und Konkretisierung des Begriffs „repository“ siehe [Heery / Andersson 2005], 2-4.

<sup>3</sup> Eine Übersicht über die aktuellen OPUS-Anwendungen gibt [http://elib.uni-stuttgart.de/opus/gemeinsame\\_suche.php](http://elib.uni-stuttgart.de/opus/gemeinsame_suche.php). Weitere Informationen zu OPUS werden unter <http://elib.uni-stuttgart.de/opus/doku/dokumentation.php?la=de> bereitgestellt.

<sup>4</sup> Basierend auf der Essener MILESS-Entwicklung. Weitere Informationen zu MyCoRe unter <http://www.mycore.de>.

ten<sup>5</sup> noch zu früh ist, die indischen Anwendungen mit den DINI-Kriterien zu messen, so kann die hier vorgelegte Analyse indischer Open Source-basierter Repositorien<sup>6</sup> doch als ein Beitrag in eine solche, qualitativ orientierte Diskussion dienen.

Im internationalen Kontext waren<sup>7</sup> und sind neben den schon genannten Produkten OPUS und MyCoRe, weiteren, national relevanten sowie den mittlerweile auch im internationalen Kontext eingesetzten Entwicklungen Fedora und CDSWare die folgenden Open Source Programme für den Aufbau digitaler Repositorien von Relevanz:

- DSpace, ein im November 2001 von MIT Libraries und Hewlett-Packard vorgestelltes Produkt, für das im Januar 2007 weltweit 190 Installationen aufgelistet werden – davon 5 in Deutschland.<sup>8</sup>
- EPrints, ein an der University of Southampton, School of Electronics and Computer Science entwickeltes Programm, das 2000 erstmalig vorgestellt wurde.<sup>9</sup> Im Januar 2007 werden 218 sog. „Archive“ mit Eprints weltweit betrieben – davon 10 in Deutschland.
- Greenstone Digital Libraries (GSDL), eine von der University of Waikato im Rahmen des New Zealand Digital Library Projekt entwickelte Open Source Software, die von der UNESCO als Anwendungsprogramm zur Distribution von digitalen Materialien sowie zum Aufbau digitaler Bibliotheken empfohlen und genutzt wird.
- Von diesen drei Open Source Produkten ist GSDL am wenigsten Workflow-orientiert; stattdessen ist es primär auf die Bereitstellung ohnehin schon vorhandener digitaler Objekte oder Metadaten ausgerichtet.<sup>10</sup>

---

<sup>5</sup> So die Aussage von Devika Madalli bei einem DINI-Workshop am 21.2.2006 im Rahmen der International Conference on Semantic Web & Digital Libraries in Bangalore, Indien (<http://drtc.isibang.ac.in/icsd/index.html>).

<sup>6</sup> Eine ausführlichere Darstellung der Analyse und ihrer Ergebnisse wird in einer gesonderten, englischsprachigen Publikation erscheinen.

<sup>7</sup> Einen allerdings z. T. veralteten funktionalen Überblick gibt [OSI 2004].

<sup>8</sup> Allgemein zu Dspace: [http://wiki.dspace.org/index.php/EndUserFaq#About\\_DSpace](http://wiki.dspace.org/index.php/EndUserFaq#About_DSpace). Siehe auch <http://wiki.dspace.org/index.php//DspaceInstances>

<sup>9</sup> Vgl. für weitere Informationen zu EPrints <http://www.eprints.org> sowie [Harnard / Tansley 2000].

<sup>10</sup> Greenstone gibt auf seiner Homepage zahlreiche Beispielsanwendungen der Software. Die Anzahl der berichteten Downloads betrug nach Auskunft des Entwicklers bislang mehrere 10.000. Die Anzahl der Implementierungen in Deutschland ist nicht bekannt, dürfte aber sehr gering sein.

Alle drei Programme sind Unix-basiert. GSDL läuft allerdings auch auf Windows-Betriebssystemvarianten. Es ermöglicht zudem den Export und die Bereitstellung der Ressourcen der Repositorien z. B. auf eine CD-ROM und seine Anwendungsoberfläche ist in einer Vielzahl von Landessprachen verfügbar (vgl. <http://www.greenstone.org>).

Die Spezifika dieser Programme, insbesondere jedoch ihre Bereitstellung als Open Source Software haben dazu geführt, dass sie weltweit vorzugsweise im Hochschul- und Forschungskontext genutzt werden. Indische Einrichtungen haben nicht nur sehr frühzeitig diese Programme genutzt<sup>11</sup> und propagiert, sondern bilden – zumindest für DSpace und EPrints – mittlerweile die größte Nutzergemeinde außerhalb des anglo-amerikanischen sowie europäischen Kontextes. Irritierenderweise wurde sie bislang von den einschlägigen Untersuchungen in diesem die internationale Diskussion dominierenden regionalen Kontext weitgehend ignoriert.<sup>12</sup>

Zu den drei genannten Open Source-Programmen<sup>13</sup> werden in den einschlägigen Verzeichnissen<sup>14</sup> allein für Indien folgende Zahlen öffentlich zugänglicher digitaler Repositorien / Bibliotheken<sup>15</sup> genannt:

- *DSpace*: 22 digitale Repositorien (vgl. Anlage 1)
- *EPrints*: 7 digitale Repositorien (vgl. Anlage 2)
- *Greenstone Digital Library*: 6 Webadressen digitaler Bibliotheken mit mehreren Subbibliotheken (vgl. Anlage 3).

---

<sup>11</sup> So erfolgte z. B. die erste DSpace Anwendung außerhalb des MIT am Documentation Research & Training Centre (DRTC) des Indian Statistical Institute in Bangalore, Indien. Zudem fanden schon im Jahr 2002 Schulungsworkshops zu EPrints und DSpace in Indien statt.

<sup>12</sup> Vgl. z. B. die Untersuchung von [van Westrienen / Lynch 2005].

<sup>13</sup> Die anderen genannten Open Source-Entwicklungen – incl. Fedora – sind nach Auskunft indischer Experten bislang nicht in erwähnenswerter Weise zur Anwendung gekommen.

<sup>14</sup> Dies waren für die hier genannten Zahlen das “Registry of Open Access Repositories” (ROAR; <http://roar.eprints.org>); “Sites Powered by EPrints” (<http://www.eprints.org/software/archives/>), “DSpace installations world wide” (<http://wiki.dspace.org/index.php/DSpaceInstances>); “Digital Repositories using DSpace in India” <http://drtc.isibang.ac.in/DRTC/dspace-india.html>) und Greenstone Support Organization for India (GSOI) <http://greenstonesupport.iimk.ac.in/indiandls.htm>.

<sup>15</sup> Auffallend (und noch näher zu untersuchen) in der indischen Szene ist der Umstand, dass sehr gerne der Terminus „Digitale Bibliothek“ synonym zu digitale Repositorien benutzt wird. Diese Uneindeutigkeit spiegelt sich hier wieder.

Insbesondere hinsichtlich GSDL kann darüber hinaus von einer erheblichen Anzahl von lokalen, nicht über das WWW erreichbaren Installationen ausgegangen werden.

Ein nicht unerheblicher Teil dieser auf internationalen und nationalen Listen geführten Anwendungen war allerdings während einer vierwöchigen Untersuchungsperiode (Mitte November – Mitte Dezember 2006) nicht erreichbar. Dies korrespondiert mit Berichten in aktuellen Veröffentlichungen zur Entwicklung digitaler Bibliotheken in Indien [z. B. Ghosh/Das 2006]. Insgesamt wird dabei erkennbar, dass Fragen der Qualitätssicherung und Erreichbarkeit der Anwendungen von entscheidender Bedeutung sind:

- „The uniform resource locator (URL) has been changed in due course of time, but other web pages link to the earlier URL (i.e., dead link).
- The web server is not operational on 24/7 timeframe that makes it inaccessible to the scholars across the world. ...
- The service has been discontinued but the web pages that are linked to this IR are not updated.
- The web server is down for the time being.” [Das/Sen/Dutta 2005, 96]

Darüber hinaus wurde festgestellt (ebd. 94), dass nach anfänglicher Euphorie der Zugang nachgewiesener Dokumente z. T. deutlich zurückging oder gar ganz nachließ. Dass solche Probleme nicht nur Indien-spezifisch sind, zeigen jüngste Veröffentlichungen in der internationalen Fachliteratur (vgl.[Gierveld 2006]; [Sale 2006]).

Vor diesem Hintergrund ergibt sich mindestens die folgende Frage: Was sind die speziellen Rahmenbedingungen und Probleme, unter denen diese Repositorien in Indien eingerichtet und betrieben werden?

Deren Beantwortung kann interessante Hinweise auf die Entwicklung digitaler Bibliotheken in einem der Länder Südasiens mit hoher wirtschaftlicher und wissenschaftlicher Dynamik geben. Darüber hinaus können Überlegungen abgeleitet werden, in welcher Weise und unter welchen Voraussetzungen dort die von DINI formulierten Qualitätsanforderungen zum Tragen kommen können.

## **2 Zielsetzungen und Methodik**

Ziel dieser Untersuchung ist die Ermittlung von Stand und Rahmenbedingungen des Aufbaus und Betriebs digitaler, öffentlich über das Internet zugänglicher, mit-

tels Open Source Software betriebener Repositorien in Indien. Wie dargelegt, sollen auf dieser Grundlage nicht nur Erkenntnisse über und ggf. Empfehlungen zu den landesspezifischen Rahmenbedingungen des Betriebs solcher Anwendungen abgeleitet werden, sondern gleichzeitig auch allgemeine Hinweise, die für die Betreiber digitaler Bibliotheken – im Sinne institutioneller und fachlicher Repositorien – von Interesse sein könnten.

Die Studie ist weitgehend empirisch-analytisch angelegt, indem

- die ermittelten Anwendungen hinsichtlich ihres Angebots nach einfachen Kennzahlen sowie ergänzenden Angaben analysiert werden;
- auf der Grundlage einer Fragebogen-basierten Umfrage Selbstauskünfte zu den Entstehungs- und Bereitstellungsbedingungen der Repositorien ermittelt und in das Gesamtbild einbezogen werden.<sup>16</sup>

Aus fachlichen wie auch politischen Gründen wurde hierbei bewusst darauf verzichtet, in diesem Stadium der Entwicklung die in Deutschland entwickelten DINI-Kriterien an die indischen Repositorien anzulegen.

### 3 Öffentlich zugängliche digitale Repositorien in Indien: Analyse und Befragung

#### 3.1 Analyse der Repositorien

##### 3.1.1 Organisations- und softwarebezogene Ergebnisse

**Verfügbarkeit.** Von den insgesamt 35 (D 22 / E 7 / G 6)<sup>17</sup> mit den drei genannten Open Source-Programmen betriebenen digitalen Repositorien waren während des Untersuchungszeitraums 10 (D 8 / E 1 / G 1), also knapp 30% permanent nicht erreichbar. Wesentliche Gründe hierfür dürften Netzwerk- und Serverprobleme sein.<sup>18</sup>

---

<sup>16</sup> Die Untersuchung erfolgte Mitte November bis Mitte Januar 2007 und fand im Rahmen eines Forschungsaufenthaltes des Autors in Indien statt. Der Fragebogen wird vom Autor auf Nachfrage gerne zur Verfügung gestellt. Logistische und technische Unterstützung erfolgte durch das Documentation Research & Training Centre (DRTC), Indian Statistical Institute, Bangalore Centre (<http://drtc.isibang.ac.in/DRTC/index.html>).

<sup>17</sup> D steht für DSpace-, E für EPrints- und G für GSDL-Anwendungen.

<sup>18</sup> Darüber hinaus deutet diese Diskrepanz darauf hin, dass entweder die aktualisierende Kommunikation zwischen den Anwendern und den Anbietern der Übersichtslisten nicht funktioniert oder die Werbefunktion dieser Listen dominiert.

**Hauptsächliche Anwender.** Die Hauptgruppe der Anwender der öffentlich zugänglichen digitalen Repositorien sind bislang wissenschaftliche Forschungseinrichtungen von nationaler Bedeutung wie Indian Institute of {Science / Technology / Management ...}, National Institute of {...} etc. sowie einige wenige Universitäten oder andere wissenschaftliche Einrichtungen. Sie sind mit den internationalen Entwicklungen im Bereich digitaler Bibliotheken vertraut und verfügen darüber hinaus zumeist über interne Strukturen (IT-Know-how; Netzanbindung; 24/7-Unterstützung im technischen Bereich), die für die Einrichtung von digitalen Repositorien notwendige Voraussetzung sind.

**Genutzte Open Source Software.** Auch wenn bedacht werden muss, dass die öffentlich verfügbaren Listen kein vollständiges Bild der Softwarenutzung für digitale Repositorien geben, so ist dennoch offensichtlich, dass bislang mehrheitlich DSpace als Anwendungsprogramm genutzt wird.

Fachveröffentlichungen zur Entwicklung der indischen Bibliotheksszene (z. B. [Jain/Babbar 2006]) machen außerdem deutlich, dass Open Source-Software in diesem Segment dominiert und proprietäre, d.h. herstellerspezifische Software bislang keine Marktrelevanz hat.

**Funktionale Unterschiede zwischen den genutzten Open Source-Programmen.** Zielsetzung und Funktionsweise der beiden Programme DSpace und EPrints unterscheiden sich nachgewiesenermaßen nur geringfügig (vgl. z. B. Madalli 2003, Borghoff et al. 2005 sowie Madalli/Seth 2005). Beide sind originär auf die Archivierung von Materialien durch Wissenschaftler bzw. Mitglieder der entsprechenden Zielgruppe, den damit verbundenen Workflow sowie die Recherche und Bereitstellung der archivierten Ressourcen ausgerichtet. Auch wenn mit ihnen ebenso die Bereitstellung von retrospektiv digitalisierten Ressourcen möglich ist, so war dies nicht ihr originärer Entwicklungszweck.

Im Gegensatz dazu wurde Greenstone Digital Library primär dafür entwickelt, schon digital vorhandene Materialien bzw. Metadaten recherchefähig zusammen- und bereitzustellen. Die Ausgabe und Distribution auf digitalen Datenträgern ist insbesondere für datentechnisch noch unzulänglich vernetzte Anwendungszusammenhänge von wesentlicher Bedeutung.<sup>19</sup> Workflow-bezogene Funktionen fehlen. Hingegen sind die Funktionen der Oberflächengestaltung incl. der Bereitstellung in

---

<sup>19</sup> Bezeichnenderweise schlägt sich dies nicht in einer höheren Anzahl von öffentlich verfügbar gemachten Anwendungen nieder, da GSDL vorzugsweise als Einzelplatzversion oder im Intranet genutzt wird.

einer Vielzahl von Landessprachen<sup>20</sup> sowie die anwendungsspezifische Rechercheaufbereitung weit entwickelt. Greenstone zielt insofern mit seiner anderen Funktionalität auf primär bereitstellungs- und distributionsorientierte Vorhaben. Der Input wird bei Greenstone normalerweise von Informationsspezialisten vorgenommen. Die andersartige, faktisch breiter gefächerte Anwendergemeinschaft von GSDL ergibt sich auch durch die Verfügbarkeit auf Windows-Betriebssystemen, womit ggf. auch weniger IT-erfahrene Anwendergruppen erreicht werden.

**Unterstützung in Indien.** Für alle drei Anwendungsprogramme gibt es generelle und speziell auf Indien ausgerichtete Unterstützungsangebote. Noch am wenigstens ausgeprägt ist dies für **EPrints**, für dessen Interessenten zuletzt 2004 in Indien mehrere Workshops durchgeführt wurden, deren Teilnehmer wohl die Multiplikatorenfunktion noch nicht im erwarteten Maße realisiert haben. Für die Anwender besteht die Möglichkeit, an der weltweiten email- und Forumskommunikation von EPrints-Anwendern teilzuhaben.

Weitaus stärker, personell präsenter und aktiver ist die Unterstützung für **DSpace** und seine Anwender durch das DRTC des Indian Statistical Institute, das auch einen Vertreter in den DSpace-Beirat entsendet. Mittels bislang 12 Schulungsworkshops in Indien, der Bereitstellung von die Installation vereinfachenden Skripten sowie einer Mailingliste wurden und werden Anwender unterstützt.

Für **GSDL** wurde im Jahre 2006 eine Unterstützerguppe für den Raum Südostasien gegründet<sup>21</sup>, die seitens der Hauptentwickler von GSDL in Neuseeland unterstützt wurde und wird. Die weit über die sechs aufgelisteten Anwendungen hinausgehende Nutzergruppe wird primär mittels einer Mailingliste beraten. Durch die UNESCO-Kooperation ist GSDL insbesondere in bibliotheks- und informationsorientierten Fachkreisen bekannt und wird dort angewendet.

In allen drei Fällen sind die Unterstützungsaktivitäten wesentlich initiiert und getragen von internationalen, persönlichen Kontakte indischer Informationswissenschaftler, die – finanziert durch nationale oder internationale Förderprogramme – Gelegenheit zu fachlichem Austausch in einem jener Länder hatten, die die fachwissenschaftliche repository-Diskussion dominieren (vgl. FN 12).

---

<sup>20</sup> GSDL ist in Englisch, Französisch, Spanisch und Russisch verfügbar; für die Anwendungsoberfläche wird eine Vielzahl von weitere Sprachen – darunter mehrere indische Sprachen – angeboten.

<sup>21</sup> Für weitere Details siehe <http://greenstonesupport.iimk.ac.in/>.



### 3.1.2 Statistische Analyse

Der Großteil der Repositorien / digitalen Bibliotheken wurde in den Jahren 2004-2006 eingerichtet. U. a. vor diesem Hintergrund ist erklärbar, dass die meisten Sammlungen bislang nur aus mehreren Hundert Dokumenten bestehen. In diesem Punkt folgen sie offensichtlich sehr ähnlichen Entwicklungsstadien wie die Repositorien in anderen Ländern.<sup>22</sup>

Für die erreichbar gewesenen Datenbestände ergibt sich danach folgendes Bild: Tabelle 1 zeigt die Anzahl der bereitgestellten Dokumente in DSpace-Anwendungen, Tabelle 2 die in Eprints-Anwendungen und Tabelle 3 die in den öffentlich zugänglichen GSDL-Bibliotheken in Indien. Die Darstellung in Tabelle 3 ist abweichend, um die Daten für die GSDL-charakteristischen „Bibliotheken“ (Teilkollektionen von Dokumenten u./o. Metadaten) wiedergeben zu können. Die dort z. T. sehr hohen Zahlen ergeben sich durch den Nachweis von (importierten) bibliografischen Metadaten (MD). Die Datensätze sind ggf. mit dem Volltext oder einer digitalen Kopie eines Dokumentes verlinkt. Berechnungen auf der Grundlage der in den Repositorien gemachten Jahreszuordnungen bei der Eingabe der Datensätze haben für die meisten der untersuchten Repositorien eine jährliche Inputrate von unter 100 Dokumenten ergeben.

Bei den gegebenen Zahlen zeigt die bei DSpace wie auch EPrints verfügbare Option der Differenzierung des Dokumentbestandes in sog. „communities“ und „collections“ (DSpace-Terminologie für Ober- und Untergruppen von Dokumenten) noch wenig erkennbare Vorteile – auch wenn sie für die Wissenschaftler zur Eingabe ihrer Dokumente in der für sie jeweils fachlich relevant Rubrik vorbereitet wurden.

Daten zu DSpace-basierten Repositorien	D 1	D 2	D 3	D 5	D 6	D 7	D 8	D 11	D 12	D 14	D 15	D 16	D 17	D 20
Anzahl von Dokumenten im Repitorium	202	84	120	1197	175	193	1867	199	257	357	477	324	1415	146
Anzahl der Autoren (z.T. geschätzt)	312	75	154	788	82	192	2305	192	199	451	673	256	590	39
Anzahl der „collections / communities“	5	1	4	8	19	6	4	4	1	1	9	1	8	8

Tabelle 1: DSpace-basierte Repositorien: Basisdaten (November 2006)<sup>23</sup>

<sup>22</sup> Vgl. die Angaben bei [van Westrienen / Lynch 2005] sowie die Zahlen bei [Heery / Andersson 2005] S. 21f, die für das Jahr 2005 auf gleichem Niveau liegen.

<sup>23</sup> Die Nummern der Tabelleneinträge (z. B. D5) korrespondieren mit der Nummerierung der in den Anlagen aufgelisteten Repositorien.

Daten zu EPrints-basierten Repositorien	E 1	E 2	E 4	E 5	E 6	E 7
Anzahl von Quellen im Repitorium	153	170	5836	2038	1376	75
Anzahl der Autoren (z.T. geschätzt)	200	85	300	960	*	100
Anzahl der „collections / communities“	9	18	9	11	*	10

Tabelle 2: EPrints-basierte Repositorien: Basisdaten (November 2006); \* keine Angaben

Daten zu Greenstone-basierten Repositorien	G 1	G 2	G 4	G 5	G 6
Anzahl von Dokumenten / Metadatensätzen im Repitorium bzw. in seinen sog. Bibliotheken (z.T. geschätzt)	~ 200 ~ 200 ~ 100 ~ 40 ~ 130 ~ 340 55 4 ~390 10	72 >17000 MD 28 7 58 7 57 ~ 100 ~ 1000 13	>3000  keine Angaben	12 36 7 15 20 7 5 7 210 10 15 6 9 4 13 1	40000 MD
Anzahl der Autoren	Angaben kaum ermittelbar				
Anzahl der „Libraries“	10	10	2	16	1

Tabelle 3: GSDL-basierte Bibliotheken: Anzahl der in den jeweiligen „Bibliotheken“ bereitgestellten Dokumente bzw. Metadatensätze (November 2006)

- Nur wenige der Anwendungen weichen von der Basiskonfiguration (sog. „out of the box“-Konfiguration) der Programme ab, so dass kollektionsbezogene Funkti-

onserweiterungen sowie institutionsbezogene Designvarianten zumeist nicht zum Tragen kommen. Entsprechend spröde ist das Erscheinungsbild der meisten Anwendungen. Problematischer im Hinblick auf die Integration der nachgewiesenen Dokumentbestände ist jedoch der Umstand, dass die Repositorien in den meisten Fällen

- nicht durch klar gekennzeichnete Links, in vielen Fällen sogar überhaupt nicht über die Homepage der jeweiligen Bibliothek/Informationseinrichtung der Institution aufrufbar sind;
- funktionale stand-alone-Anwendungen geblieben sind, so dass normalerweise keine bestandsübergreifende Recherchen in gedruckten und digital bereitgestellten Ressourcen möglich ist. Auch die Integration in die sonstigen bibliothekarischen Anwendungsprogramme fehlt.

Größtenteils sind die Repositorien als institutionelle Sammlungen angelegt. In einigen Fällen ist dies auch in einer einfachen Sammelleitlinie formuliert. Detaillierter und elaborierter sind größtenteils Aufgabendarstellungen und Sammelleitlinien, die für die themenorientierten Repositorien angelegt wurden. Es sind dies vorzugsweise jene, die auch bei OAIster (<http://oaister.umdl.umich.edu/o/oaister/>) für den Austausch von Metadaten registriert sind bzw. einen OAI-basierten Datenaustausch realisieren (z. B. LDL, Librarians' Digital Library, <https://drtc.isibang.ac.in/>).

### **3.2 Fragebogenbasierte Umfrage zum Betrieb der digitalen Repositorien**

#### *3.2.1 Inhaltliche Zielsetzung, Verteilung und Rücklauf*

Parallel zur direkten Analyse der Websites und Repositorien erfolgte im November / Dezember 2006 eine Befragung der Ansprechpartner / Betreiber der ermittelten Repositorien mittels Fragebogen.

Der Fragebogen umfasste 25 Fragen (Nr. jeweils in Klammern genannt) zu

- Grundlegenden Kontaktdaten (1-2)
- Typ und Inhalt des digitalen Repositoriums (3-5)
- Ablauf und Voraussetzungen für die Einrichtung des digitalen Repositoriums (6-11)
- den Umständen, unter denen das DR betrieben wird (12-19)
- der genutzten Open Source Software (20-21)
- Verfügbarkeit und Nutzung des digitalen Repositoriums (22-24)
- Zukünftigen Erwartungen bezüglich des digitalen Repositoriums (25)

Nach einem Pretest im Rahmen direkter Befragung wurde der Fragebogen am 23.11.2006 an 30 Einrichtungen (D 19 / E 5 / G 6)<sup>24</sup> mittels e-mail verschickt. Innerhalb 16 Tagen wurde ein Rücklauf von 13 beantworteten Fragebögen verzeichnet (D 9 / E 2 / G 2). Mit der für derartige Umfragen überdurchschnittlichen Rücklaufquote von 43,3% können die nachfolgenden Aussagen auf einer vergleichsweise stabilen Grundlage erfolgen.

### 3.2.2 Ergebnisse der Umfrage

Die Auswertung der Antworten auf die Fragebögen zeigen die nachfolgend zusammengefassten Trends und Ergebnisse:

- Für die untersuchten Repositorien lässt sich ein klarer Trend zu institutionellen Repositorien erkennen.
- Die meisten der Repositorien bieten bislang textbasierte Materialien an.
- Mangels Input durch Autoren aus den Institutionen überwiegen bislang digitalisierte Dokumente in den Repositorien. Dennoch haben originär digitale Materialien schon einen hohen Anteil an den Beiträgen.
- Die meisten der Repositorien wurden im Jahr 2005 begonnen, nur einige wenige früher in den Jahren 2001/2002 bzw. später im Jahr 2006.
- Sie wurden zumeist auf Initiative einer institutionellen Bibliothek bzw. eines/r Bibliothekars/in, in einigen Fällen auch nach Initiative externer Aktivisten, kaum aber von den Wissenschaftlern oder der Verwaltung der wissenschaftlichen Einrichtungen realisiert.
- Kenntnisse über die Einrichtung eines digitalen Repositoriums kamen zumeist von Dritten wie z. B. den Initiatoren und Referenten der Schulungsworkshops. Sie brachten ihr Fachwissen bei der eigentlichen Einrichtung des Servers mit ein. Alternativ dazu erfolgte dies durch die Spezialisten aus den IT-Abteilungen.
- In ungefähr der Hälfte der Fälle gibt es weder detaillierte Sammelleitlinien noch sog. *institutional policies* zum Aufbau der Repositorien. Nur in Einzelfällen (z. B. National Institute of Technology, Rourkela) wird eine institutionelle Selbstverpflichtung formuliert und im Kontext des Repositoriums bekannt gemacht. Allerdings stellen die meisten der Trägerorganisationen die notwendige Infrastruktur in Form von Hardware, Netzwerkanbindung und Arbeitszeit von Fachpersonal zur Verfügung.

---

<sup>24</sup> Mangels Verfügbarkeit der Website bzw. der Angabe eines Ansprechpartners konnte nur an soviel Repositorienverantwortliche ein Fragebogen verschickt werden.

- Die wöchentliche Arbeitszeit für den Betrieb des Repositoriums wird mit 3 bis maximal 10 Stunden angegeben.
- Die folgenden technischen Probleme werden genannt, die Einfluss auf den Betrieb nehmen:
  - Serverprobleme
  - „übergelaufene“ Logfiles
  - Unmöglichkeit, den gesamten Inhalt des Repositoriums zu exportieren
  - Netzwerkprobleme
  - Formatprobleme bei den angelieferten Dokumenten.<sup>25</sup>
- Die folgenden organisatorischen Probleme wurden genannt:
  - fehlende Bereitschaft der Autoren ihre Veröffentlichungen selbst ein
  - zuliefern bzw. der Sammlung zur Verfügung zu stellen
  - fehlende Erlaubnis der Verleger, Dokumente auf diesem Wege bereit zu stellen.
- Bibliothekare überprüfen die von den Autoren eingegebenen Metadaten (immanentes Workflow-Feature bei DSpace und Eprints) – sofern sie nicht ohnehin selbst diese Daten eingeben. Der damit verbundene arbeitszeitliche Aufwand variiert in Abhängigkeit von der Anzahl der aufgenommenen Dokumente von einer Arbeitswoche für eine Person pro Monat bis hin zu zwei Vollzeitmitarbeitern.
- Die meisten der Anwender scheinen mit der von ihnen genutzten Software zufrieden zu sein<sup>26</sup>, auch wenn einige deutlich machten, dass sie noch erhebliche Erfahrungsdefizite bei der Nutzung sehen. Einige Anwender haben mit allen 3 Produkten Erfahrungen gesammelt, um die für ihre Bedürfnisse geeignetste Software zu ermitteln.<sup>27</sup>
- Die monatliche Zugriffsrage ist mit 20 bis zu 6-7000 Zugriffen sehr unterschiedlich und spiegelt die sehr heterogene Anwendungslandschaft wider. Hierzu passt

---

<sup>25</sup> Die genannten Probleme könnten auch auf eine noch nicht hinreichende Kompetenz im Umgang mit der Software hindeuten.

<sup>26</sup> Aufgrund der statistischen Gegebenheiten bei der Beantwortung des Fragebogens bezieht sich diese Aussage vorrangig auf die DSpace-Anwender.

<sup>27</sup> Daraus haben sich dann auch einige wenige nicht weitergeführte Anwendungen ergeben, die das Gesamtbild verfälschen. Das produktspezifische Know-how scheint häufig an Einzelpersonen gebunden zu sein und sich allein schon deshalb einem produktbezogenen Vergleich zu entziehen. Selten wurden bislang wirkliche Produktvergleiche wie bei [Maddali / Seth 2005] auf der Grundlage von Tests veröffentlicht – überraschenderweise auch nicht bei den Asian Digital Libraries-Konferenzen.

auch die Aussage, dass einige Betreiber diese Zahl nicht wissen und bislang auch nicht zu ermitteln versuchten.

- Die öffentliche Bereitstellung von statistischen Daten wird bislang nur von ganz wenigen Betreibern praktiziert.
- Nur einige wenige der Sammlungen sind nicht nach dem Konzept 24/7 verfügbar. Diesbezüglich muss aber einbezogen werden, dass die Antwortenden ohnehin eine positive Auswahl darstellen.
- Die Erwartungen an die zukünftige Entwicklung der Sammlung zielen insbesondere darauf, die Bereitschaft der Wissenschaftler zur Weitergabe ihrer Veröffentlichungen zu steigern und / oder die Anzahl der bereitgestellten Dokumente – und damit auch die Attraktivität der Sammlungen – durch retrospektive Digitalisierung zu erhöhen.

### **3.3 Zusammenfassung der statistischen Analyse und der Umfrageergebnisse**

Die Ergebnisse der Angebotsanalyse wie auch der Umfrage geben ein komplementäres Bild von der aktuellen Situation der mittels den drei Open Source Software-Produkten bereitgestellten, öffentlich zugänglichen digitalen Repositorien in Indien. Danach wurden – abgesehen von einer kleinen Anzahl – die meisten der Sammlungen erst in den vergangenen 2 Jahren angelegt und beinhalten demzufolge noch vergleichsweise wenige Dokumente (s. o. FN 22). Als besonders erfolgreich erweisen sich zwei Typen von Repositorien:

- Jene institutionellen Repositorien, die klar formulierte und offensichtlich auch organisationsintern kommunizierte Sammelrichtlinien entwickelt haben und hierfür ggf. auch mehr als nur infrastrukturelle Unterstützung seitens der Organisation(leitung) erfahren.
- Jene, die als themenorientierte Repositorien einen klar definierten fachlichen Sammel- bzw. Bereitstellungsauftrag formuliert haben.

## **4 Zusammenfassung und Schlussfolgerungen**

Die Studie hat gezeigt, dass viele international orientierte indische Wissenschaftseinrichtungen sehr aktiv bei der Einrichtung von digitalen Repositorien sind. Insbesondere dort sind auch in den Bibliotheks- und Informationsabteilungen Spezialisten mit internationalen Kontakten aktiv. Dass dabei zu den in der internationalen Open Source- bzw. Open Access-Szene gängigen Produkten gegriffen wurde und wird, ist vermutlich auf eine Mischung aus Sprach- und Kulturorientierung sowie finanzielle bzw. logistische Unterstützungen zurückzuführen. Zudem funktio-

niert diese Szene sehr ausgeprägt über persönliche Beziehungen, so dass entweder persönliches bzw. institutionelles Engagement alleine oder in Verbindung mit strukturpolitischer Einflussnahme zur Multiplikationseffekten führt(e), als deren Ergebnis zahlenmäßig wachsende Anwendergemeinden entstanden. Angesichts der Einwohnerzahl und der Entwicklungsdynamik des Schwellenlandes Indien ist diese Anzahl im internationalen Vergleich zwar schon hoch, liegt aber noch weit unter dem tatsächlichen Potenzial und Bedarf des Landes.

Allerdings haben nur einige dieser Anwendungen einen professionellen Stand erreicht, der erwarten lässt, dass sie auch längerfristig erfolgreich existieren werden.

Vorzugsweise dort wo der anwendungstechnische Enthusiasmus Einzelner oder unterstützende Infrastruktur aus einer IT-Abteilung besteht, wurden die Anwendungen bislang im Erscheinungsbild und funktional über die Basisinstallation hinaus entwickelt. Die tatsächlichen technischen Kenntnisse zur Betreuung, Einbindung in andere Anwendungen und zur funktionalen Weiterentwicklung der konkreten Software-Installation sind in den meisten Fällen auch aus Anwenderperspektive noch nicht so ausgeprägt, dass die verfügbaren Anwendungen mehrheitlich als Vorzeigebispiele angesehen werden können. Hier scheint noch erheblicher Schulungs- und weiterführender Betreuungsaufwand notwendig zu sein, um ein konsistenteres Gesamtbild in diesem Bereich zu erzeugen. Schulungen allein für die Initialisierung der Einrichtung eines Repositoriums sind sicher ein wichtiger Schritt, ihnen müssen jedoch weiterführende, anwendungsbegleitende Unterstützungsmaßnahmen folgen, die über die Beantwortung von Fragen via e-Mail-Liste hinausgehen.

In jenen Fällen, in denen Anwendungen erfolgreich realisiert wurden und auch auf einen zielgruppenspezifischen Bedarf stoßen, erfahren sie eine eindrucksvoll hohe Nachfrage und reziprok dann auch um so mehr innerorganisatorische Unterstützung. Ein sehr gutes Beispiel hierfür ist das für den Nachweis und die Bereitstellung von Forschungspublikationen angelegte institutionelle Repositorium ePrints@IISc (<http://eprints.iisc.ernet.in/>) des Indian Institute of Science in Bangalore.

Insofern lässt sich die eingangs formulierte Frage nach den speziellen Rahmenbedingungen und Problemen, unter denen Repositorien in Indien eingerichtet und betrieben werden, weitgehend mit den auch international geltenden Erkenntnissen (vgl. z. B. [Gierveld 2006]) beantworten – sieht man einmal von regionalen kommunikationstechnischen Einschränkungen ab.

Als zentrale Erfolgsfaktoren sind daher zu nennen:

- Technische Kompetenz bei den verantwortlichen Betreuenden und/oder entsprechende Unterstützung innerhalb der Organisation bzw. von außen

- Verankerung der Sammel- und Bereitstellungsidee bei der relevanten Zielgruppe sowie den Führungskräften der Organisation bzw. den peers des Faches
- Sammelrichtlinien und -zielsetzungen, die in Form von mission oder policy statements formuliert und propagiert sowie von der Organisationsleitung unterstützt werden.
- Personelle und organisatorische Infrastrukturunterstützung seitens der zentralen Informationseinrichtung, im Normalfall also der Bibliothek oder einer entsprechenden fachlichen Informationsabteilung.

Den initialisierenden Qualifizierungsangeboten zur Einrichtung der Repositorien müssen deshalb weiterführende, betreuende Angebote folgen, um die Nachhaltigkeit und Qualität der Sammelaktivitäten abzusichern. Dem muss durch entsprechende Basisqualifizierung in der Ausbildung von Bibliothekaren / Informationsfachleuten sicher Rechnung getragen werden, es besteht aber darüber hinaus erheblicher Weiterbildungs- und Betreuungsbedarf, der vermutlich durch die Aktivisten der Open Source-Szene nicht geleistet werden kann. Alternativ zu der von den Anwendern von Open Source-Produkten häufig erwarteten entgeltfreien Beratung aus dem Kollektiv der software- und anwendungsbezogenen Know-how-Träger könnte sich insofern zumindest für kostendeckende, ggf. sogar gewinnorientierte Beratungs- und Betreuungsdienstleistungen ein Marktsegment entwickeln. Deren Rahmenbedingungen und Chancen sind allerdings nicht nur für Indien ein noch weitgehend unerforschtes Thema. Solche Dienstleistungen könnten für die Ausbildung von Qualitätsanforderungen orientiert am DINI-Zertifizierungskonzept sicher von Vorteil sein.

Die für Indien identifizierten Qualitätsanforderungen korrespondieren bezeichnenderweise zumindest mit den ersten drei der von DINI aufgestellten Zertifizierungsanforderungen (vgl. Einführung). Sie sind eindeutig produktspezifisch und dürften auch für die Anwendungen in Deutschland zutreffen. Schließlich sind auch in Deutschland bislang erst 19<sup>28</sup> dieser mindestens 84 Publikationsserver von DINI zertifiziert.

## 5 Literatur

Borghoff et al. (2005). Vergleich bestehender Archivierungssysteme, München 2005. Hrsg. von nestor, Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland (Nestor materialien ; 3) URN: *urn:nbn:de:0008-20050117016*

---

<sup>28</sup> <http://www.dini.de/dini/zertifikat/zertifiziert.php>; Stand Januar 2007.



- Das, Anup Kumar; Sen, B K, Dutta, Chaitali (2005). Collection Development in Digital Information Repositories in India. – Vishwabharat@TDIL; 2005; 17; 91-96, <http://hdl.handle.net/1849/203>
- DINI-Arbeitsgruppe „Elektronisches Publizieren“ (2007): DINI-Zertifikat Dokumenten- und Publikationsservice 2007. (DINI Schriften 3-de) [Version 2.0, September 2006]; <http://edoc.hu-berlin.de/series/dini-schriften/2006-3/PDF/3.pdf>
- Harnard, Stevan; Tansley, Robert (2000). Eprints.org Software for Creating Institutional and Individual Open Archives. D-Lib Magazine, 6 (2000) 10, <http://www.dlib.org/dlib/october00/10inbrief.html#HARNAD>
- Heery, Rachel; Andersson, Sheila (2005). Digital Repositories Review. Report for JISC, 2005; [http://www.jisc.ac.uk/uploaded\\_documents/digital-repositories-review-2005.pdf](http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf)
- Ghosh, S: B.; Das, Anup Kumar (2006). Open access and institutional repositories – a developing country perspective: a case study of India. World Library and Information Congress: 72nd IFLA General Conference and Council “Libraries: Dynamic Engines for the Knowledge and Information Society” 20-24 August 2006, Seoul, Korea, [http://www.ifla.org/IV/ifla72/papers/157-Ghosh\\_Das-en.pdf](http://www.ifla.org/IV/ifla72/papers/157-Ghosh_Das-en.pdf)
- Gierveld, Heleen (2006). Considering a Marketing and Communications Approach for an Institutional Repository. – In: Ariadne 49, October 2006 <http://www.ariadne.ac.uk/issue49/gierveld/>
- Jain, P.K., Babbar, Parveen (2006). Digital libraries initiatives in India. In: The International Information & Library Review 38 (2006), 161-169
- Madalli, Devika (2003). A Digital Library of Library and Information Science using Dspace. – In: DRTC Workshop on Semantic Web, 8<sup>th</sup>-10<sup>th</sup> December, 2003, DRTC Bangalore, 7. Paper G.
- Madalli, Devika; Seth, Renu (2005). Open Source Software for Building Digital Repositories. – In: Kaul, H. K., Sen, Gayathri (Ed.): Library and Information Networking, Proceeding of the National Convention on Library and Information Networking (NAC-LIN 2005) held at PES Institute of Technolog, Banglore, August 22-25, 2005 , 394-405
- Open Society Institute (2004). Open Society Institute: A Guide to Institutional Repository Software, 3rd Edition, New York, August 2004; [http://www.soros.org/openaccess/pdf/OSI\\_Guide\\_to\\_IR\\_Software\\_v3.pdf](http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf)
- Oßwald, Achim (2006). Deutsch-Indisches E-Learning-Projekt zum Thema Digitale Bibliotheken. In: Bibliotheksdienst 40 (2006), 7, 903-905
- Sale, Arthur (2006). The Patchwork Mandate. 2006, <http://eprints.utas.edu.au/410/>
- van Westrienen; Lynch, Clifford A. (2005) Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005, D-Lib Magazine, 11(2005) 9; <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>

## **6 Anlagen**

### **6.1 Anlage 1: Digitale Repositorien mit DSpace**

Central Plantation Crops Research Institute, Kasargod <http://210.212.229.11:8080/dspace/>

GB Pant University of Agriculture & Technology, Pant Nagar <http://202.141.116.205/dspace/>

Indira Gandhi Institute for Development Research, Mumbai <http://oii.igidr.ac.in:8888/dspace>  
INFLIBNET, Ahmedabad (nicht erreichbar) <http://dspace.inflibnet.ac.in/>  
Indian Institute of Astrophysics Repository <http://prints.iia.res.in/browse-title>  
Indian Institute of Management, Kozhikode <http://dspace.iimk.ac.in/>  
Indian Institute of Science, Bangalore – Electronic Theses and Dissertations of Indian Institute of Science <http://etd.ncsi.iisc.ernet.in/>  
Indian Institute of Technology, New Delhi <http://eprint.iitd.ac.in/dspace/>  
Indian Institute of Technology, Kharagpur (nur Intranet Zugang) <http://10.17.32.208/dspace>  
Indian National Science Academy, New Delhi (nicht erreichbar)  
<http://drtc.isibang.ac.in/insa>  
Indian Statistical Institute, Bangalore <http://library.isibang.ac.in:8080/dspace/>  
LDL: Librarians' Digital Library, DRTC <https://drtc.isibang.ac.in/>  
National Centre for Radio Astrophysics, Pune (nicht erreichbar)  
<http://ncralib.ncra.tifr.res.in/dspace>  
DSpace at National Chemical Laboratory, Pune, India <http://dspace.ncl.res.in/dspace/index.jsp>  
National Institute of Oceanography, Goa <http://drs.nio.org/>  
National Institute of Technology, Rourkela <http://dspace.nitrkl.ac.in/dspace/>  
Raman Research Institute, Bangalore <http://dspace.rrri.res.in/>  
Sri Venkateswara University, Tirupati (nicht erreichbar) <http://202.41.85.207:8080/dspace>  
University of Hyderabad, Hyderabad (nicht erreichbar) <http://202.41.85.207:8080/dspace>  
ICFAI BUSINESS SCHOOL (IBS), Ahmedabad <http://202.131.96.59:8080/dspace/>  
Bangalore Management Academy, India (nicht erreichbar) <http://59.92.116.53:8080/dspace>  
Vidyanidhi Digital Library & E-Scholarship Portal, University of Hyderabad (India);  
<http://dspace.vidyanidhi.org:8180/dspace/> (nicht erreichbar)

## 6.2 Anlage 2: Digitale Repositorien mit EPrint

DU Eprint Archive (156); <http://eprints.du.ac.in/>  
ePrints@IIMK: Indian Institute of Management Kozhikode Scholarship Repository (?);  
<http://eprints.iimk.ac.in/>  
Indian Institute of Information Technology (?); <http://eprints.iiita.ac.in/>; (nicht erreichbar)  
Indian Institute of Science, Bangalore, India (5751); <http://eprints.iisc.ernet.in/>  
NAL-IR (2182); <http://nal-ir.nal.res.in/>  
OpenMED@NIC (1293); <http://openmed.nic.in/>  
OneWorld South Asia, Open Archive Initiative <http://open.ekduniya.net/>

## 6.3 Anlage 3: Digitale Bibliotheken mit Greenstone Digital Library

Archives of Indian Labour <http://www.indialabourarchives.org/>  
Digital Library Collections Built at NCSI <http://144.16.72.147/gddl/cgi-bin/library>

Developmental Informatics Laboratory <http://www.mlasia.iitb.ac.in/gsd/cgi-bin/library> (nicht erreichbar)

Electronic Theses and Dissertations IIT Bombay <http://www.library.iitb.ac.in/%7Emnj/gsd/cgi-bin/library>

Indian Institute of Management Kozhikode <http://www.iimk.ac.in/gsd/cgi-bin/library>

Indian Institute of Science Publications Database Bangalore

PRABHAVI: Indian Institute of Science Publications Database

<http://vidya-mapak.ncsi.iisc.ernet.in/cgi-bin/library?a=p&p=home>



# The Modality Principle in Multimedia Learning

## An Open Question: When Speech Fails to Foster Learning?\*

*Klaus D. Stiller*

University of Regensburg

D-93040 Regensburg

*klaus.stiller@psychologie.uni-regensburg.de*

### Abstract

Learning is fostered when spoken rather than written texts are used concurrently with dynamic or static pictures in instructional settings. This is mainly caused by split attention and can be explained by the cognitive theory of multimedia learning (Mayer, 2001) or the cognitive load theory (Sweller, 1999). Numerous studies have provided evidence for modality effects (Ginns, 2005), but recently, studies also revealed a vanishing or reversal of modality effects in favor of visual texts. This can be attributed to learner characteristics or instructional design, but knowledge about relevant features is still scarce.

### 1 Multimedia learning

In general, learning is defined as a relatively sustainable change of a subject's behavior or potential for a behavior related to a certain situation, which is created by repeated experiences of subjects in particular situations. Changes can occur as regards cognitive, psychomotor, affective-motivational or social aspects (Lukesch, 1998). On the other hand, teaching refers to an intended and planned initiation and fostering of learning in order to achieve teaching goals. Mostly, it is focused on the cognitive goals of learning and instruction.

Learning and instruction can be realized by using media. The most prominent modes for presenting information are written and spoken texts as well as static and dynamic pictures. Most medial presentations involve at least one of the verbal and

---

\* Published in: Osswald, Achim; Stempfhuber, Maximilian; Wolff, Christian (eds.) (2007). Open Innovation. Proc. 10. International Symposium for Information Science. Constance: UVK, 129-144.

one of the pictorial modes. In Anglo-American literature, learning concurrently with texts and pictures is known as multimedia learning (Mayer, 2001). That is important to note, because in most German literature the word *multimedia* is linked to computer programs, indicating that these programs feature at least a minimum of interactivity and a presentation of information using verbal and pictorial presentation codes, partly in a dynamic way, for the auditory and visual senses (Issing & Klimsa, 2002).

One of the prominent theories which describe and explain meaningful learning with texts and pictures is the cognitive theory of multimedia learning (CTML; Mayer, 2001). It is based on the dual channel assumption, the limited capacity assumption and the active learning assumption. The dual channel assumption proposes that humans process pictorial and verbal information in two separate systems, a visual/pictorial and an auditory/verbal system (Fig. 1). The systems are functionally independent of each other, but may interact. The processing of pictures occurs mainly in the visual/pictorial channel and that of spoken words in the auditory/verbal channel. Written words are initially processed in the visual/pictorial channel before they move to the auditory/verbal channel in the working memory. Furthermore, it is possible for verbal and pictorial information to cross the channel barriers. Verbal information that can be (easily) imagined or pictorial information that can be (easily) verbalized may find access to the other channel.

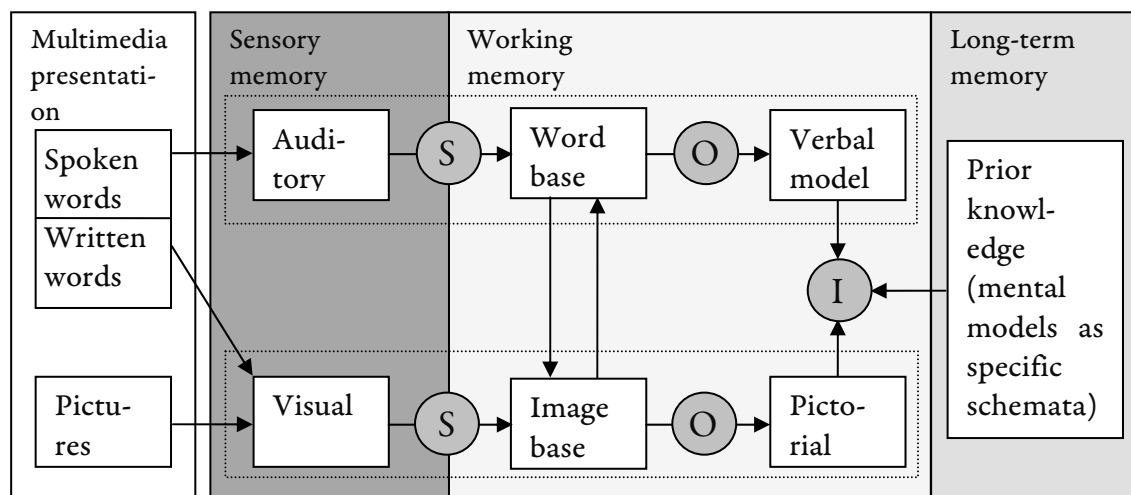


Figure 1: Framework for the CTML (Mayer, 2001). S stands for selecting information, O for organizing information into pictorial and verbal models, and I for integrating mental models and prior knowledge.

The limited capacity assumption suggests that the information processing capacity in each of the two channels is limited. Within capacity limits the learner has to ac-

comply with all the information processing necessary for (meaningful) learning. Especially the cognitive load theory (CLT) emphasizes the role of the working memory in learning, and the problem of occupying working memory by useful and unnecessary processes (Sweller, 1999). A working memory with limited capacity only allows processing of 2 to 4 elements (an estimation by Sweller, 1999), which is less than the magical number 7 plus/minus 2 according to Miller (1956). This is due to the fact that processing information requires more load than merely retaining items in working memory (Sweller, 1999). Additionally, working memory is also limited in time. If rehearsal of memory items is not possible, items fade within seconds. Due to these severe limitations of our working memory, learning is often difficult, because tasks are complex. Learners have to retain a large number of elements in their minds and relate them or search for actions that bring them nearer to solving a problem. This often exceeds the working memory capacity. Working memory resources are divided into various parts that are covered by special kinds of cognitive processes which in turn exert loads on working memory. These loads can be detrimental to or beneficial for learning. The components of cognitive load are extraneous, germane and intrinsic load, which are intrinsically linked to the source of the load. Intrinsic load is created by task complexity and is mainly needed for understanding. Germane load goes beyond understanding and is bound to the learning processes in a real sense – processing information in order to build mental representations and storing these representations. Extraneous load results from the presentation manner of the material and is often the main source that is detrimental to learning. As long as extraneous, intrinsic and germane load do not occupy the entire working memory capacity, some memory capacity that could be used for processing is still available. Mayer and Moreno (2002, p. 108) regard the CLT as “an important component in any theory of learning that is intended to guide the design of multimedia learning environments”. Accordingly, they regard the CLT as an important component of the CTML.

The active learning assumption states that learners have to process information actively in order to achieve meaningful learning. Active cognitive processing includes the selection of relevant information from words and pictures, the mental organization of this information to form coherent verbal and pictorial representations, and the mental integration of these representations with one another and with prior knowledge from long-term memory. Selection and organization take place in separate channels (see the dual channel assumption) with integrative processing bridging the channels. The SOI processes “generally do not occur in a rigid linear order, but rather in an iterative fashion” (Mayer, 2001, p. 130).

To a great extent, learning depends on presentational features, which determine the amount of extraneous load. As extraneous load is detrimental to learning, designers should strive towards minimizing it. Modality effect is one of the effects connected with and caused by extraneous load. The modality principle in instructional design is directly linked to modality effect and the associated empirical evidence, which are dealt with in the next section.

## **2 The modality principle**

The modality principle recommends using spoken rather than written texts in conjunction with pictures, because learners learn more effectively in computer-based and book-based environments when spoken rather than visual texts are used concurrently with dynamic or static pictures. The superiority of spoken over written texts was demonstrated in multiple studies, largely using low prior knowledge learners (Ginns, 2005). Modality effects were observed in various measures: (1) less mental effort (a measure for cognitive load) with spoken words while learning (e. g. Tabbers, 2002; Tindall-Ford, Chandler, & Sweller, 1997), (2) less time on problem-solving tasks (e. g. Jeung, Chandler, & Sweller, 1997; Mousavi, Low, & Sweller, 1995), (3) higher scores on various tasks of retention, matching verbal and pictorial information, transfer and practice (e. g. Kalyuga, Chandler, & Sweller, 1999, 2000; Moreno & Mayer, 1999), and (4) shorter reaction time in a secondary task (e. g. Brünken, Plass, & Leutner, 2004; Brünken, Steinbacher, Plass, & Leutner, 2002).

Modality effect is thought to be composed of two parts: visual/pictorial processing channel overload and split-attention effect. Firstly, when words and pictures are presented visually, the words are processed at least initially in the visual/pictorial channel, allowing this channel to become overloaded, the visual/auditory channel remaining unused. Thus, written texts exert extra load on working memory, because they initially use the visual processing channel before the information is transmitted to the verbal processing channel. On the other hand, auditory texts enter the auditory/verbal processing channel directly, thus releasing the visual/pictorial channel from unnecessary load. The working memory resources gained might be allocated to germane load. Overall, the interchannel transition of visual text information is often believed to be an automatic process and that it therefore does not occupy many resources (e. g. Tabbers, 2002; Schmidt-Weigand, 2006).

Secondly, a split-attention effect occurs, because attention has to be spread between visual text and pictures, which results in less relevant material being selected for further processing (especially with system-paced instructions including dynamic pres-



entations). The split-attention effect always occurs when learners have to divide their attention between different sources of information that must be integrated mentally to achieve comprehension (Sweller, 1999). A central characteristic of the separated sources of information is that all sources are needed in order to reach full understanding. The multiple sources of information are related to one another and are unintelligible in isolation (Sweller, 1999). The sources of information can be separated in time or space. If learners have to split their attention between such sources of information, extraneous load on working memory will increase. This means, for example, that learners must search for corresponding information from the disparate sources and keep it concurrently in working memory in order to enable integration processes. This exerts a heavy extraneous load on working memory. If the increased extraneous load results in working memory being overloaded, this in turn hinders learning. An auditory presentation of text reduces these unnecessary processes; no splitting of visual attention between pictures and visual text is necessary in order to select relevant information and subsequently to organize and integrate it in working memory. A simultaneous audio-visual presentation increases the likelihood of learners being able to retain corresponding words and pictures in working memory at the same time. Thus it is also more likely that the learners engage in active meaningful learning, in particular in integrative processing of verbal and pictorial information.

Optimal temporal and spatial contiguity is achieved with spoken words. Verbal and pictorial information can be selected in parallel and integrated immediately. Spoken texts release the learner from extraneous load e. g. from extended visual search and accompanying processes as well as from actively retaining information in working memory longer than necessary.

## **2.1 Overview of studies confirming modality effects on performance**

According to Tabbers (2002), older studies which support modality effect are limited concerning subject matter, length of study time, settings and pacing of presentation. The implemented instructions contained contents taken from the exact sciences such as geometry, mathematics, physics and electrical sciences, and they mostly did not last longer than 5 minutes. They were conducted in laboratory settings and the pacing of presentations was determined by the computer system or the procedure in non-computer settings. Since then, however, empirical evidence of modality effect has also been found with learner-paced or learner-sequenced instructions consisting of up to 90 minutes' learning time in classroom and laboratory settings, sometimes using non-scientific contents (e. g. Brünken, Seufert, & Zander,

2005; Moreno & Mayer, 2002; Mayer, Dow, & Mayer, 2003; Moreno, Mayer, Spires, & Lester, 2001).

Mousavi et al. (1995) confirmed the modality effect on the basis of worked-out examples concerning geometry problems in a book-based environment. Students looked at a printed diagram and listened to or read an explanation. Geometry problems which were similar to the ones used in the instructional phase were solved faster when students had listened to the narration; geometry transfer problems were also solved faster, but the statistical test was not significant. Kalyuga et al. (1999, 2000) found equivalent results as regards transfer performance using diagrams and verbal explanations in computerized environments on the topic of soldering and manufacturing.

Tindall-Ford et al. (1997) have shown with book-based instructions about electric circuits that modality effect only holds for material of high, but not of low complexity. With highly complex material, an overload of working memory was forced back within the limits of capacity by using auditory texts; with material of simple or medium complexity, although the degree of cognitive load changed when using visual or auditory texts, the load stayed within the limits of working memory capacities. Leahy, Chandler, and Sweller (2003) found equivalent effects. They used questions tapping low or high element interactivity knowledge and only revealed a modality effect with the latter.

Modality effect was also evidenced with computerized instructions lasting between 10 and 20 minutes on scientific (human circulatory system) as well as non-scientific contents (the historic city of Florence) that were system-paced in laboratory settings (Brünken & Leutner, 2001; Brünken et al., 2002, 2004) or learner-paced in classroom settings involving pupils and university students (Brünken et al., 2005). Learners using spoken texts showed a faster reaction time in a secondary task, understood texts and pictures better and had more success at solving transfer problems.

Craig, Gholson, and Driscoll (2002), Mayer and Moreno (1998), and Moreno and Mayer (1999) have shown that modality effect holds for retention, matching (learners had to place verbal labels on the corresponding parts of pictures) and transfer performance in computer-based environments which used explanations about lightning or brakes, using animation and narration or on-screen text. The on-screen text was placed below the animation (Mayer & Moreno, 1998; Moreno & Mayer, 1999) or when a pedagogical agent was used in a speech bubble placed near the top left-hand corner of the screen (Craig et al., 2002). Learners with narrated anima-

tions could remember or recognize more of the key ideas and were better at matching pictorial and verbal information and using the knowledge gained for solving new problems.

Schmidt-Weigand (2006) conducted four experiments using animation with spoken or visual texts about lightning formation. In his first experiment (Exp. 1 in Ch. 2), he found modality effects for retention and transfer both when visual text was placed below the animation or near to the pictorial part in question. Furthermore, in two experiments (Exp. 2 in Ch. 2 and Exp. 1 in Ch. 3) using animation and static pictures, only weak evidence was found for modality effects on verbal retention, transfer and visual recall (drawing task). Transfer performance was not influenced by modality in either of the two experiments, retention was better with narrated animation and visual recall was mainly better with narrated static pictures in Experiment 2 in Chapter 2. In the fourth experiment (Exp. 2 in Ch. 3), he found modality effects only on visual knowledge; no effects on retention and transfer were observed. Measures of overall load revealed no modality effect throughout all experiments.

Mayer et al. (2003, Exp. 1) used an interactive, self-explanatory environment using a cartoon agent. The instruction showed a picture of an electric motor; the five parts of the electric motor were clickable. When a part was clicked on, it was highlighted and a list of thirteen questions was shown at the top right-hand corner of the screen. After clicking on a question, the answer was presented either as spoken text or written text, which was also shown in the top right-hand corner of the screen. Sometimes, the verbal answer was accompanied by an animation. The information units could be revisited or replayed. Study time took about 20 minutes. Modality effect was shown to hold for transfer in an interactive computer-based environment.

Moreno et al. (2001) as well as Moreno and Mayer (2002) presented an agent-based multimedia game on botany. Learners had to design eight and five plants respectively, whereby each plant should flourish on a different planet characterized by its own climate. An agent introduced task and planet and gave advice concerning the learners' choices of roots, stems and leaves. The agent spoke or showed written messages. The game could be paced by learners; therefore depending on the number of plants, it took between 24 and 28 or between 10 and 16 minutes, respectively, to complete the game. In five experiments, a modality effect of retention and transfer was found.

Using a computerized simulation environment, O'Neil et al. (2000) staged a virtual training episode on an aircraft fuel system. Learners had to grasp the structure of a

fuel system and how the system functions. Therefore, learners were placed in a virtual environment via head-mounted display, and the instructional system provided a guided exploration of the fuel system in a virtual world. Trainees could navigate around and through the internal components of the fuel system by using a 3D-mouse and could even conduct a series of tests. Auditory and visual text groups learned for approximately 26 to 28 minutes on average. A modality effect on retention was not found, but was supported as regards matching and transfer performance. The knowledge structure assessed by concept maps tended also to be better with spoken texts.

Stiller (2007, Exp. 1) investigated modality effects and their dependence on prior knowledge with learner-paced and learner-sequenced instructions under fixed study time. Instructions were hierarchically structured hypermedia systems on the physiology of the visual system. Texts were accessed by clicking on parts of pictures. University students learned with visual or auditory texts for 25 minutes. A modality effect on structural knowledge as measured by completing pictures by drawing and a verbal description task were found in favor of auditory texts; retention, transfer and labeling pictures were not affected. Modality effects did not depend on prior knowledge.

Stiller (2007, Exp. 4) investigated modality effects and their dependence on attitude towards computers with hierarchically structured hypermedia systems concerning the physiology of the perceptual systems. Subjects studied three learner-paced and learner-sequenced instructions under learner-controlled study time (86 min. on average), whereby detailed texts were presented visually, auditorily and bimodally. Introductions, summaries and glossary were always presented visually. Concerning performance at verbal retention, labeling of representational pictures and a line diagram, there is evidence for modality effect with negative-attitude learners but not with positive-attitude learners. Additionally, a pure modality effect occurred as regards labeling mnemonic pictures. Spoken text prolonged learning time, this mainly being due to an intensive use of the glossary as regards the auditory group. In a follow-up study (Stiller, 2007, Exp. 5), modality effects and their dependence on prior knowledge were investigated. The same instructions as in Exp. 4 were used, but the glossary was eliminated. University students studied with two instructions for 69 minutes on average, whereby detailed texts were presented visually or auditorily. A modality effect was found for the labeling of mnemonic and representational pictures as well as transfer; verbal retention and labeling of a line diagram were not affected. Concerning transfer performance, low-prior-knowledge learners benefited from auditory texts and high-prior-knowledge learners could learn equally well with

visual or auditory texts. Instructional usage was only marginally influenced by text modality and prior knowledge.

Rinck and Glowalla (1996) found no modality effect on performance with longer-lasting and learner-paced instructions (5 lessons each lasting approx. 2 hours), but they found advantages in the speed of answering performance tasks by learners who had been instructed with auditory texts.

## **2.2 Overview of studies showing contradictory evidence**

It seems very easy to generalize the findings to other learning environments like interactive media, but this is more complicated than one would suppose, in view of the empirical evidence. Tabbers (2002) and Stiller (2007) realized a number of experiments in order to investigate modality effect in newly applied situations. But they could only partly support a modality effect; moreover, they revealed some contradictory results. Schmidt-Weigand (2006) also showed that modality effects might even vanish when using the most demanding presentation modes consisting of animation and verbal explanation. Modality effects also proved to be dependent on learner features such as prior knowledge and computer attitude (Stiller, 2007).

Schmidt-Weigand (2006; Ch. 4) showed that the modality effect disappeared when learners were able to control the pacing of a presentation. Schmidt-Weigand (2006) used explanations by animation and narration or visual text that consisted of 16 sequences presented in succession; the learners could decide when to start each sequence. It was predicted that modality effects would vanish with learner-paced instructions. This was confirmed for retention, visual knowledge and transfer; performance differences were insignificant. Mental effort measures also showed no influence of text modality. Study time did not influence performance. Learner-paced presentation durations did not differ when using auditory or visual texts (on average 183 seconds,  $SD = 42.86$ ).

Tabbers' (2002) aim was to generalize modality effect to longer instructions from a different content domain in an ecologically valid classroom setting. He used web-based multimedia lessons which lasted from 26 to 70 minutes and dealt with a didactical model of instructional design, focusing especially on learner-pacing. The results of the four experiments can be summarized as follows. (1) Modality effect could be widely replicated for mental effort. Spoken words put less load on working memory with both system-paced and learner-paced instructions. (2) Modality effect of retention and transfer could be partly replicated with system-paced instructions. (3) Performance increased and modality effect vanished with more time

on task when using system-paced instructions. (4) Modality effect vanished or more often reversed for retention and transfer with learner-paced instructions.

In Experiment 1, static diagrams were explained by auditory or visual texts. Referenced diagram parts were highlighted in red. Cueing of pictorial parts was introduced to minimize visual search and thus extraneous load. Learners controlled the pace of instruction. Information units were put in linear order and learners could move forward and backward along this line. Mental effort tended to be lower with spoken words during instruction, was significantly higher for the visual text group when solving retention tasks and similar for treatment groups as regards solving transfer tasks. Retention and transfer performance was better after students had received the instruction with visual texts. Thus a modality reversal effect for retention and transfer was found, which suggests a visual presentation of texts with learner-paced instructions.

In Experiment 2, static diagrams were accompanied by narrations or on-screen texts, but this time instructions were system-paced. Referenced diagram parts were cued to minimize visual search. Study time was 26.2 minutes. The participants, students from a teacher training college, were considered to have no prior knowledge. Study time, mental effort during instruction and testing as well as retention and transfer performance were assessed. With spoken words, less mental effort was reported, but no differences on retention or transfer were found between groups. The audio-text group achieved the same performance level as the visual-text group, but with less mental effort.

In Experiment 3, modality effects were investigated in dependence on instructional pacing. Instructions consisted of static diagrams accompanied by narrations or visual texts and paced by system or learner. Referenced diagram parts were colored in red to minimize visual search. The system-paced instruction took 26.2 minutes; the learner-paced instruction was not restricted in time. Information units were arranged in linear order and learners could only move forward in line with the sequence. The university students participating were considered to have no prior knowledge of the topic. Study time, mental effort during instruction and testing as well as retention and transfer performance were assessed. No modality effect on study time was found. On average, the self-paced audio group learned only 2.6 minutes less (28.3 min.) than the self-paced visual group (30.9 min.). The average study times of the self-paced groups were not much longer than the duration of the system-paced instructions (26.2 min.). Mental effort was not influenced by text modality. Modality effects for retention and transfer occurred with system-paced instructions; with learner-paced instructions no effects were observed at all. Addi-

tionally, transfer performance was influenced by pacing. The learner-paced groups outperformed the system-paced groups, mainly as a result of the lower performance of the system-paced group that had to learn with visual texts. The transfer performances of the other three groups had the same level.

The aim of Tabbers' (2002) fourth experiment was to test the idea of time on task as the cause for the vanishing of modality effect with learner-paced instructions. Six conditions were used, resulting from the combination of two modality and three pacing conditions. The modality conditions consisted of static diagrams accompanied by (1) narrations or (2) visual texts, which were paced (1) by system as every narration part was played once, (2) by system as every narration part was played twice, or (3) by learner. Information units were put in linear order. Study time was 19.3 minutes for single system-paced instructions and 38.6 minutes for double system-paced instructions. The university students participating were considered to have no prior knowledge of the topic. Study time, mental effort during instruction and testing as well as retention and transfer performance were assessed. The average study time duration of the learner-paced instruction was 24.1 minutes. There was no difference in average study time between learner-paced instructions using auditory or those using visual texts. Mental effort was lower for students learning with spoken texts, no effects on mental effort were observed for solving retention and transfer tasks. Retention and transfer performance were affected by modality and pacing. A modality reversal effect of retention and transfer occurred with learner-paced instructions, whereas no modality effects appeared for system-paced conditions. Moreover, the double system-paced groups performed better at retention tasks than the single system-paced groups.

Stiller (2007, Exp. 2) investigated the effects of prior knowledge and computer attitude on modality effect. 10<sup>th</sup> grade German grammar school pupils learned with visual or auditory texts on the human eye in learner-paced and learner-sequenced instructions for 13 minutes. Modality effect was found to depend on prior knowledge and computer attitude. Overall, the low-prior-knowledge and negative-attitude learners in the auditory and visual group, respectively, did not differ significantly, but the high-prior-knowledge and positive-attitude learners benefited most from visual texts and were handicapped by spoken texts. Transfer and drawing performance were not affected by modality, but a superiority of visual texts over auditory texts was found as regards verbal retention and labeling pictures. The instructional usage was only marginally influenced by prior knowledge and text modality.

### 3 Discussion

Modality effect was found and amply evidenced by using system-paced instructions for experiments. For system-paced instructions which determine and limit time on presented information, spoken texts are regarded to be more effective because texts and pictures are perceived simultaneously; there is no danger of missing information or increased cognitive load due to split-attention. Learners studying a system-paced instruction based on visual texts suffer from dividing attention between texts and pictures, thus experiencing more extraneous load and learning disruption. Moreover, the visual entry channel might also be overloaded.

Tabbers (2002) argued that the standard rationale for explaining modality effect which is offered by the CLT and the CTML cannot explain the vanishing or the reversal of modality effect. As modality effect is assumed to result from split-attention and visual processing channel overload, these factors must be essentially influenced in order to obtain contradictory results. Tabbers (2002) rejects the assumption that presenting visual text and picture at the same time easily overloads the visual processing channel. Learners must split their attention between visual text and picture, but visual text is immediately transformed to phonological code and does not hinder the subsequent processing of pictorial information in the visual processing system. So, what are instances of factors influencing modality effect?

To a large extent, modality effect depends on medial features. One of them is the pace of the presentation, which determines the time available for meaningful information processing. Tabbers (2002) and Schmidt-Weigand (2006) have shown that extended time on task compensates mainly for split-attention effects, thus preventing working memory from being overloaded or at least relieving working memory from extraneous load. Extended time on task enables learners to spend more processing time on the verbal and pictorial information presented. Hence learners are more likely to select relevant information and to integrate it adequately within extended time. Thereby compensation of mainly split attention is balanced out by increased mental effort (Tabbers, 2002). To sum up, time on task can account for the vanishing of modality effect with system-paced and learner-paced instructions.

The argument of time on task cannot explain modality reversal effects that only occurred with interactive instructions. Time on task can only explain a vanishing of modality effect due to a compensation of split attention. The reversal of modality effect can be explained by an increased flexibility of information processing. Therefore Tabbers (2002) proposed that reading visual texts is a more active process that allows regulation of information processing. For this reason, learner-paced in-



structions with visual texts might be more appropriate in certain situations, suiting the learners' individual, cognitive needs. It is much easier, for example, to jump back and forth in written text in order to access or revisit pieces of information; contrarily, it is quite difficult to do this within audio files. A more flexible work-through might lead to a greater chance of retaining corresponding verbal and pictorial information in working memory, of improving mental representations by noticing hitherto unnoticed, but essential information, of mapping verbal and pictorial models adequately and of using prior knowledge for knowledge construction. In this context, listening is a rather passive process which makes it especially suitable for linear presentations with minimal interactivity, as in system-controlled instructions. Learner-pacing sometimes does not go along with normal listening behavior. Therefore spoken texts are sometimes not suitable for learner-pacing. To sum up, learner-pacing in the sense of guaranteeing sufficient time on the task in question only suggests that spoken and visual texts should be equally effective, because the factors detrimental to learning with visual texts can be compensated; learner-pacing in the sense of guaranteeing more flexible processing can account for the reversal of modality effect.

Modality effects also depend on learner characteristics. This was shown by Stiller (2007, Exp. 2) who found a vanishing of modality effect with pupils who have low prior knowledge and negative computer attitude, and a reversal of modality effect with high-prior-knowledge and positive-attitude pupils. But Stiller (2007, Exp. 1) also found modality effects with university students, independent of prior knowledge, using comparable learner-paced instructions as in Exp. 2. Hence, it seems to depend on the learners as to how they make use of learner-pacing as regards information processing in order to benefit from or be hindered by visual texts. Alternatively, these contradictory results might be attributed to text difficulty. Kalyuga et al. (1999) assume that speech would not be an effective instruction mode if texts were too long or complex, as this might also overload working memory. The possibility of overload is attributed to the fleeting character of auditory information that makes it more difficult to retain information in the memory. Visual text is permanent and can therefore be referred to repeatedly. The texts might have been more difficult for pupils than for students, hence pupils might have experienced overload with auditory texts. Overall, text difficulty might not be a cause of vanishing or reversal of modality effects in Tabbers' (2002) experiments, because the texts used were kept concise.

Jeung et al. (1997) believe that the reduction of working memory load with visual-auditory presentations would only enhance learning if mental resources were not

devoted to an extensive, visual-based search in order to coordinate auditory and visual information. They found auditory texts to be helpful when a diagram was easy to process, but not when it was difficult to process. Their results showed that a modality effect was not observed with highly complex pictures; auditory text was only beneficial with complex pictures when visual attention was guided by visual aids like the flashing of pictorial parts concurrently with the corresponding text presentation. As Tabbers (2002) and Stiller (2007) cued the relevant pictorial parts of the pictures, this is not regarded as a cause for a vanishing of modality effects. In addition, the learning materials used were not of low complexity, so this could neither be a cause for a vanishing of modality effects (Tindall-Ford et al., 1997; Leahy et al., 2003).

As we have seen, it is not easy to predict the circumstances under which it is better to use visual texts with interactive instructions, because there is also ample evidence of modality effect with such instructions, with effects seeming to depend on learner characteristics. Hence, future research must specify which characteristics of instructions and learners determine, whether a modality effect will occur, vanish or reverse. Against that background, the decisions between auditory and visual text productions could be made more easily.

## 4 References

- Brünken, R., & Leutner, D. (2001). Aufmerksamkeitsverteilung oder Aufmerksamkeitsfokussierung? Empirische Ergebnisse zur "Split-Attention-Hypothese" beim Lernen mit Multimedia. *Unterrichtswissenschaft*, 29, 357-366.
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual task methodology: Auditory load and modality effects. *Instructional Science*, 32, 115-132.
- Brünken, R., Seufert, T., & Zander, S. (2005). Förderung der Kohärenzbildung beim Lernen mit multiplen Repräsentationen. *Zeitschrift für Pädagogische Psychologie*, 19, 61-75.
- Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual task methodology. *Experimental Psychology*, 49, 109-119.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62, 233-246.
- Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94, 428-434.
- Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15, 313-331.

- Issing, L. J., & Klimsa, P. (Hrsg.). (2002). *Information und Lernen mit Multimedia und Internet* (3. vollständig überarbeitete Auflage). Weinheim: Psychologie Verlags Union.
- Jeung, H. J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, 17, 329-343.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351-372.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, 126-136.
- Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology*, 17, 401-418.
- Lukesch, H. (1998). Einführung in die pädagogisch-psychologische Diagnostik. Regensburg: Roderer.
- Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press.
- Mayer, R. E., Dow, G., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology*, 95, 806-813.
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90, 312-320.
- Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, 12, 107-119.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91, 358-368.
- Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, 94, 598-610.
- Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based multimedia learning: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, 19, 177-214.
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319-334.
- O'Neil, H. F., Mayer, R. E., Herl, H. E., Niemi, C., Olin, K., & Thurman, R. A. (2000). Instructional strategies for virtual aviation training environments. In H. F. O'Neil & D. H. Andrew (Eds.), *Aircrew training and assessment* (pp. 105-130). Mahwah, NJ: Erlbaum.
- Rinck, M., & Glowalla, U. (1996). Die multimediale Darstellung quantitativer Daten. *Zeitschrift für Psychologie*, 204, 383-399.
- Schmidt-Weigand, F. (2006). Dynamic visualizations in multimedia learning: The influence of verbal explanations on visual attention, cognitive load and learning outcome. Retrieved 1-31-2006, from <http://geb.uni-giessen.de/geb/volltexte/2006/2699/>.

- Stiller, K. D. (2007). Computerised multimedia learning. Modes of text presentation and access to text. Hamburg: Verlag Dr. Kovač.
- Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, AU: ACER Press.
- Tabbers, H. K. (2002). The modality of text in multimedia instructions. Refining the design guidelines. Heerlen: Open University of the Netherlands.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When to sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3, 257-287.

# Cultural Differences in Human Computer Interaction: Results from Two Online Surveys\*

*Rüdiger Heimgärtner*

Siemens AG, SV I IS RD ADI  
Siemens VDO, Siemensstraße 12, 93055 Regensburg, Germany  
{ruediger.heimgaertner}@siemens.com

## **Abstract**

This paper describes a method to obtain quantitatively discriminating cultural interaction indicators and their values for cross-cultural Human Computer Interaction (HCI) design as preparatory work for cross-cultural adaptive navigation and multi-media systems. The method has been implemented in a tool for cross-cultural HCI analysis. Two online studies temporally displaced by one year using this tool, regarding cultural adaptability exemplified by use cases of navigation systems, revealed differences in interaction behaviour that depend on the cultural background of the users. The results will be presented and discussed to demonstrate the difficulties, but also the importance to get the cultural differences in HCI to clear the way for cultural adaptability.

## **I Determining Cultural Differences in HCI as First Step to Cultural Adaptability**

To be able to adapt navigation systems manually (adaptation) or automatically (adaptability) to the cultural needs of the user, the first step is to investigate what must be adapted, i.e. to find out the differences in the cultural needs of the users and hence the cultural differences in HCI on all levels of HCI localization (surface, functionality, and interaction). This is still one of the largest explanation gaps in cross-cultural HCI design, which has to be bridged today. Here areas like presentation of information (e. g. colours, time and date format, icons, font size) and language (e. g. font, direction of writing, naming) or dialog design (e. g. menu struc-

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10. International Symposium for Information Science. Constance: UVK, 145-157.

ture and complexity, dialog form, layout, widget positions) as well as interaction design (e. g. navigation concept, system structure, interaction path, interaction speed) are concerned (cf. [Röse et al. 01], [Heimgärtner 05]). A common approach for this is to perform qualitative and personal studies. Although this process is quite controllable, it is very expensive and time consuming. Furthermore, it is very difficult to find enough test persons to get a sample size, which enables valid application of enhanced methods of statistics. Therefore, many users have been asked online to do certain use cases to get data for studying cultural differences in HCI.

## 2 Method for Getting Cultural Differences in HCI

This section describes the background of conducting two online studies to get cultural differences in HCI: after finding potential cultural variables in HCI as well as meaningful uses cases, the test tasks, the test tool and the test setting have been developed, followed by the start of the surveys.

### 2.1 Finding Potential Cultural Variables in HCI

Hall [Hall 76] found differences in communication speed between cultures, which also imply differences in information speed (“duration of information presentation”), information density (“number of parallel pieces of information during information presentation”) and information frequency (“number of information presentations per time unit”). Using this method of literature research and analytical reasoning, more than one hundred potentially culturally sensitive variables have been identified, implemented into the “Intercultural Interaction Analysis” tool (IIA tool) and applied by measuring the interaction behaviour of the test persons with a personal computer system in relation to the culture (cf. [Heimgärtner 05]). E. g., one of the variables is measuring the acceptance of the “life-like” character “Merlin”.<sup>1</sup> According to Prendinger and Ishizuka (cf. [Prendinger et al. 04]), such avatars can reduce stress during interaction with the user. Hence, the agent “Merlin” was implemented in the IIA tool to offer his help every 30 seconds. On the one hand, according to cultural dimensions, which describe the behaviour of human beings of different cultures, like high uncertainty avoidance or high task orientedness, it was expected that German users switch off the avatar very soon (compared to Chinese users), because they do fear uncertain situations (cf. [Hofstede et al. 05]). Furthermore, they do not like to be distracted from achieving the current task (cf. [Halpin

---

<sup>1</sup> The virtual assistant „Merlin“ is part of the interactive help system of Microsoft Office™.

et al. 57]). On the other hand, if applying the cultural dimension of face saving, it should be the other way around. If Chinese users make use of help very often, they would loose their face (cf. [Victor 97]). The test with the IIA tool was designed to help to reveal the empirical truth to such questions amongst other things (cf. [Heimgärtner 05]).

## 2.2 Finding Use Cases and Test Tasks

The most interesting use cases possess a high degree of interactionality. In order to limit the scope of research, representative and demonstrative use cases have been restricted for cross-cultural human machine interaction (HMI) in automotive navigation systems (cf. [Heimgärtner 05]). One such significant use case is e. g. map display. What map direction is best according to the user's cognitive style? How many points of interest (POI) should be presented to the user? A hypothesis like "there is a high correlation of high information density to relationship-oriented cultures such as China" should be confirmable by adjusting more POI by Chinese users compared to German users. So, the use case "map display" was simulated by the *map display test task* to measure the number of pieces of information on the map display regarding information density (e. g. restaurants, streets, POI, etc.) (Figure 1).

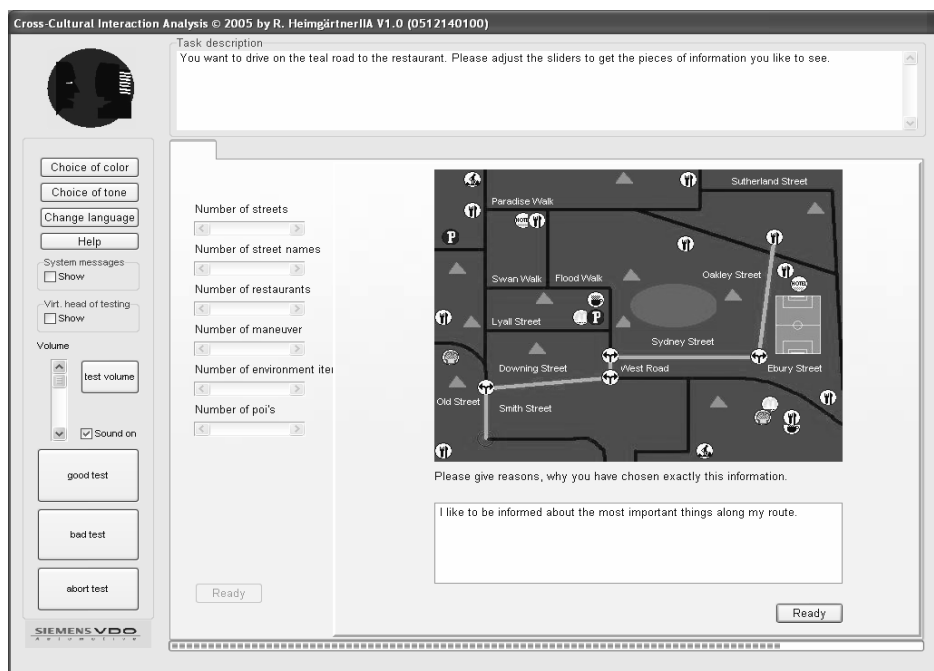


Figure 1: Screenshot of the "map display test task" during the test session with the IIA data collection module. The user can define the amount of information in the map display by adjusting the scroll bars.

Based on this principle, this test tool can also be used to investigate the values of other cultural variables like widget positions, menu structure, layout structure, in-

teraction speed, speed of information input, dialog structure, etc. Every one of the test tasks serves to investigate other cultural aspects of HCI. E. g., the special use case “manoeuvre guidance” has been implemented into the *manoeuvre guidance test task*, where the test user has to adjust the number and the time distance of the manoeuvre advice messages on the screen concerning frequency and speed of information. The test tasks (use cases) are localized but designed semantically identical for all users: participants of many different cultures can do the test. The collected data is partly quantitative (related to all test persons, e. g. like the mean of a Likert scale) and partly qualitative (related to one single test person, e. g. answering open questions) (cf. [De la Cruz et al. 05]). Moreover, the collected data sets have standard format so that anyone can perform own statistical analyses. This also means that the results of this study are verifiable because they can be reproduced using the IIA tool.

### 2.3 Test Setting with the IIA Tool

A user test session with the IIA tool comprises five parts: collection of demographic data, test tasks, VSM94 questionnaire, evaluation of results by the user, and debriefing questionnaire. The demographic questionnaire delivers knowledge about the cultural background of the user (like mother tongue, languages, nationality, and residence in foreign countries). The developed and implemented test tasks in the IIA tool serve to motivate the user to interact with the computer and to test hypotheses. To analyze the cultural attitudes of the users, the value survey module (VSM94) has to be filled in by the user (cf. [Hofstede 94]). The VSM94 contains 26 questions to determine the values of the cultural dimensions using the indices from Hofstede that characterize the cultural behaviour of the user (cf. [Hofstede 91]). The results of the VSM94 and of the test tasks are presented to the user who then has to estimate whether or not the cultural and informational values found correlate or match to him. The debriefing part reveals the purpose of the test to the user in detail and collects data regarding the usability of the test system, the perceived difficulty of the test in general as well as if the user has recognized the implemented hypotheses in the test tasks. During the whole test session, the IIA tool records the interaction between user and system, e. g. mouse moves, clicks, interaction breaks, and the values and the changing of slide bars set up by the users in order to analyze the interactional patterns of users of different cultures. All levels of the (physical, lexical, syntactic, semantic, pragmatic, and intentional) interaction model necessary for dialog design can be analyzed (cf. [Herczeg 94]).<sup>2</sup>

---

<sup>2</sup> For more details about the IIA tool and the test setting, please refer to [Heimgärtner 05] and to [Heimgärtner 06].



## 2.4 Description of the Studies

First efforts to get cultural differences in HCI scanning interaction behaviour for cultural adaptability happened in April 2005 by doing a very small local offline heuristic pre-study in Huizhou (China) and in Regensburg (Germany) with seven Chinese and eleven German students and employees of SiemensVDO. The purpose of this study was to check the usability of the IIA tool for Chinese and German users.<sup>3</sup> The two extended online studies served additionally to verify the functionality and reliability of the IIA tool and to get the preferences of users according to their cultural background. Randomly selected employees from SiemensVDO all over the world were invited per email to do the test session using the IIA data collection module by downloading it from the corporate intranet. Table 1 characterizes the two online studies regarding sample size, tests downloaded, tests aborted, valid test data sets, and return rate.

Study	Sample size	Survey period	Number of downloaded tests	Tests Aborted [%]	Number of valid test data sets	Return Rate [%]
1	600	12/14/05 – 01/14/06	166	41,5	102	16,6
2	14500	11/14/06 – 01/19/07	2803	66,8	916	6,3

Table 1: Characterization of the two online studies conducted with the IIA tool

The tests have been aborted due to the following reasons: download time too long<sup>4</sup>, no time to do the test now, test is not interesting or appealing. This type of qualitative data can help to optimize the testing equipment or to steer the direction of data analysis by asking the user for the reasons of his behaviour during the test. Only complete and valid data sets have been analyzed using the IIA data analysis module and the statistic program SPSS. The discrimination rate of classifying the users to their selected test language by the variables concerning the cultural background of the users (mother tongue, nationality, country of birth and primary residence) is 83.3% for the first and 81.9% for the second study.<sup>5</sup> Therefore, the differences in

<sup>3</sup> The IIA tool consists of three elements: a data collection module, a data analysis module and a data evaluation module. The data evaluation module serves to cross-validate the results from data analysis.

<sup>4</sup> Notably in China because of slow network connections.

<sup>5</sup> The discrimination rate has been calculated using discriminance analysis (cross validated and grouped, Wilk's Lamda in study 1:  $\lambda_{1-2}=.072^{**}$ ,  $\lambda_2=.568^{**}$ , Wilk's Lamda in study 2:  $\lambda_{1-2}=.192^{**}$ ,  $\lambda_2=.513^{**}$ ). The level of significance is referenced with asterisks in this paper (\*  $p<.05$ , \*\*  $p<.01$ ).

HCI in these studies have been analyzed in relation to three groups of test persons according to the selected test languages (Chinese (C), German (G), and English (E)) in order to reduce data analyzing costs.

### 3 Study Results

The qualitative offline pre-study, done by participative observation during and interviews after the test sessions, showed first interesting results regarding cultural dependent differences in using the IIA tool running on computer systems. There are differences between (C) and (G) concerning the order of pictures (more ordered by (G) than by (C)), test duration (longer for (C)), error clicks ((C) more than (G)) and telling the truth regarding computer experience ((C) understated their experience pretty much). In the following two online studies, some values of the implemented variables in the IIA tool show significant differences, which represent differences in user interaction according to the different cultural background of the users. Therefore, these variables can be called *cultural interaction indicators* (Table 2).<sup>6</sup>

<i>Cultural interaction indicator</i>	<i>First study</i>	<i>Second study</i>
Speed (MG)	$F(2,102)=8,857^{**}$	$\chi^2(2,916)=29,090^{**}$
MessageDistance (MG)	$F(2,102)=7,645^{**}$	$F(2,916)=16,241^{**}$
POI (MD)	$F(2,102)=3,143^*$	$\chi^2(2,916)=32,170^{**}$
MaximalOpenTasks	$\chi^2(2,102)=12,543^{**}$	$F(2,916)=15,140^{**}$
MaximalOpenTasks ratio (C,G,E)	2.5 : 1.4 : 1	1.7 : 1.03 : 1
Information speed value	$\chi^2(2,102)=17,354^{**}$	$\chi^2(2,916)=82,944^{**}$
Number of chars	$\chi^2(2,102)=16,452^{**}$	$\chi^2(2,916)=67,637^{**}$

Table 2: Cultural Interaction Indicators found in both studies

<sup>6</sup> One-way ANOVA as statistical method for comparing the means of more than two independent samples, was used to get significant cultural differences in variables, which are distributed normally. The results of the test of homogeneity of variances indicate whether ( $p > .05$ ) or not ( $p \leq .05$ ) the variables are distributed normally. A third of the potential variables was distributed normally, and hence analyzed by ANOVA. The interactional differences between the user groups separated by the test languages have been identified using the Tukey-HSD-Post-Hoc-Test after one-way ANOVA. For the remaining variables, which are not distributed normally, Kruskal-Wallis-test has been applied. The variables in the valid test data sets are not distributed comparably in the first and the second online study. Therefore, partly the same variables have been analyzed either by ANOVA or by Kruskal-Wallis-test (indicated with F or  $\chi^2$  in Table 2).

*Speed (MG)* means the driving speed of the simulated car in the manoeuvre guidance test task ((C) less than (G) and (E)). *MessageDistance (MG)* denotes the temporal distance of showing the manoeuvre advice messages in the manoeuvre guidance test task. (C) desired about 30% more pre-advice (“in x m turn right”) than (G) or (E) before turning right. This can be an indication for higher information speed and higher information density in China compared to Germany, for example. *POI (MD)* counts the number of points of interest set by the user in the map display test task. Information density increases with the number of POI and is two times higher for (C) than for (G) or (E). *MaxOpenTasks* represents the maximum number of open tasks in the working environment (i.e. running applications and icons in the Windows<sup>TM</sup> task bar) during the test session. (C) tend to work on more tasks simultaneously than (G) or (E) which can be possibly explained by the way of work planning (polychrome vs. monochrome timing, (cf. [Hall 76])) or the kind of thinking (mono-causal (sequential) vs. multi-causal (parallel) logic, (cf. [Röse et al. 01])). *Information speed value* represents the time the manoeuvre advice message is visible on the screen. (C) and (G) wanted the messages to be visible about 40% longer than (E) do. *Number of Chars* contains the number of characters entered by the user during the manoeuvre guidance and map display test tasks in answering open questions. This is explained by the fact that the Chinese language needs considerably less characters to represent words than the English or the German language.

There have also been implemented assumed cultural interaction indicators that are statistically not discriminative. In the first study, e. g. *ScrollBarChanges\_norm* ( $F(2, 102) = 0.954, p = .389$ ) shows that the number of the scrolling events triggered when moving a scroll bar slider by the user is not significantly different between the groups (C), (E) and (G).<sup>7</sup> In the second study, e. g. *TotalDialogTime* ( $F(2, 916) = 1.370, p = .255$ ) indicates that the time needed by the users to pass the dialogs of the test tasks is not significantly different between the groups.

## 4 Discussion of the Results

In this section, the disturbing variables, the classification power of the cultural interaction indicators and the reliability of the IIA tool will be discussed to argue for and to underline the plausibility of the results.

---

<sup>7</sup> The suffix “\_norm” means the value of ScrollBarChanges (= total number of scroll bar changes during the whole test session) divided by the duration of the whole test session.

#### 4.1 Disturbing Variables

If disturbing variables are known, they can be controlled in data analysis. E. g. age, gender and computer experience are variables, which can influence the results negatively. The controlled disturbing variable *Age* should not correlate with the selected test language by the user – even if it correlates only slightly. The Pearson correlation coefficients in Table 3 show that age is linearly related to test language in contrast to gender and computer experience.

<i>Pearson correlation matrix</i>	<i>First study</i>	<i>Second study</i>
Controlled disturbing variable	Test language	Test language
Test language	1,000**	1,000**
Age	0,370**	0,161**
Gender	-0,038	-0,017
Computer experience	0,174	-0,048

Table 3: Relationship between test language and controlled disturbing variables

This possible bad influence of the disturbing variable *Age* on the validity and the values of the cultural interaction indicators can be weakened by the fact that the age of the test persons of the different countries was not distributed equally in the samples. There were no Chinese test persons above the age of 40 in the first study ( $n=102$ ). The effect was lower using only test persons whose age is distributed equally in the user groups (separated by the test language) or by calculating partial correlations. This conclusion has been confirmed by the collected data of the second study: Pearson correlation and Kruskal-Wallis-test showed a lower correlation coefficient for the variable *Age* than in the first study because of  $n=916$ .

Even if computer experience is intuitively the most significant variable directly connected to interaction behaviour (e. g. interaction speed and frequency), it did not interfere significantly with the measuring process of the interaction behaviour of the users. This can be explained by the fact that computer experience was almost equally distributed in the test users at the worldwide locations of SiemensVDO because the link to the IIA tool has been sent per e-mail only to users who have Internet access and hence, who have some basic interacting experience with computers. In addition, gender does not have significant influence on the test language. Hence, in both studies, the statistical methods justified the results of the studies as correct and representative for employees of SiemensVDO: none of the controlled disturbing variables influenced the cultural interaction indicators in a way that they cannot be called *cultural interaction indicators*.

## 4.2 Classification Power of the Cultural Interaction Indicators

The cultural interaction indicators can be used to recognize the cultural interaction behaviour of the user and to relate these cultural interaction patterns to the characteristics of the user's culture. The discriminatory power of these cultural interaction indicators has been calculated using discriminance analysis. Several combinations of cultural interaction indicators contribute positively to a high discrimination rate in assigning users to their test language without knowing their nationality: only the interaction patterns within use cases or applications are known. Step-by-step discrimination analysis ("Jackknife-Method") offers iterative analysis of the best discriminating cultural interaction indicators automatically out of a given set of potential ones. The following cultural interaction indicators have been identified for the data sets of both studies exhibiting the highest classification power: interaction speed, information speed value, interaction exactness value, number of manoeuvres, ~maximal open tasks, ~POI, ~restaurants, ~streets and ~chars as well as uncertainty avoidance value. The resulting discrimination rate for classifying all test users simultaneously and correctly to their selected test languages (i.e. to the groups (C), (E) and (G)) is 60.8% for the first and 59.9% for the second study. This points to a strong similarity of the collected data as well as to the correctness of the methodology and the results of the studies (Table 4).

Study	Classification rate [%]	Test language	Predicted group membership [%]		
			Chinese	German	English
1	Cross validated total: 60,80 Wilk's $\lambda_{1-2}=.574^{**}$ , $\lambda_2=.855^{**}$ $p_{inclusion}=.05$ , $p_{exclusion}=.1$	Chinese	58,82	29,41	11,76
		German	9,09	70,45	20,45
		English	29,17	25,00	45,83
2	Cross validated total: 59,90 Wilk's $\lambda_{1-2}=.649^{**}$ , $\lambda_2=.850^{**}$ $p_{inclusion}=.05$ , $p_{exclusion}=.1$	Chinese	35,58	23,08	41,35
		German	4,55	61,76	33,69
		English	6,45	29,49	64,06

Table 4: Classification Power of the Cultural Interaction Indicators

However, the Chinese interaction behaviour is not classified very clearly in the second study (35.58%, cf. Table 4), which indicates that *in this case* the controlled disturbing variables could influence the classification power of the cultural interaction indicators. One possible explanation for this are differences in the sample sizes ( $n_C=1500$ ,  $n_E=4500$ ,  $n_G=8500$ ). Probably there are too few Chinese data sets for reasonably conducting discriminance analysis for classification to all three groups simultaneously. This fact required deeper analysis. Hence, applying the method of discriminance analysis classifying the cases into two groups (instead of three groups

at the same time), the discrimination rate increases tremendously: it goes up to 70.4% for (G) and (E) and is even 85.4% for (C) and (G). This outcome in conjunction with the weak influence of disturbing variables supports the high reliability and criteria validity of the statistical results received in the two online studies as well as the functional correctness and appropriateness of the IIA tool. Additionally, to verify the discriminating rate by a more practical method, a back propagation network has been implemented into the IIA data evaluation module. All values of the potential cultural interaction indicators of all data sets have been z-transformed and normalized to the range of [0;1] to be able to feed the input neurons with comparable data. Three output neurons indicated the test languages Chinese, German, and English. According to the network topology and learning rate, the discrimination rate climbed up to 80% for correctly classifying the users to the used test language, which also supports the correctness of the study results.

#### 4.3 Optimizing the Test Design by Intercultural Usability Engineering

The variation of the classification power of the cultural interaction indicators (cf. Table 4) can probably also be explained by different test conditions or test tasks whose design still has to be optimized applying the intercultural usability engineering process and methods even more profoundly (cf. [Honold 00]). E. g., in both studies, *NumberOfHelp* counts the number of initiations of online help by the test persons. Usually this variable was zero, which shows that help was not needed. This fact can be exploited, e. g. to indicate that the test tasks were self-explaining and comprehensible for the users. Nevertheless there are differences between the groups ((C), (G) and (E)) in using the help function ( $\chi^2(2, 916) = 1.619, p=.445$ , ratio (C:G:E) = 5.6:1:1.4). This can possibly explained by the fact that a German designer developed the IIA test (I). Hence, the German imprinted design as well as the explanation of the test tasks shall be optimized for Chinese users in future tests.

#### 4.4 General Cultural Interaction Indicators

The results of the two online studies show that *HCI between the Chinese, German, and English-speaking participants differs significantly*. A possible explanation of this fact is probably grounded in subconscious cultural differences imprinted by primary culture and learning the mother tongue, which leads to different HCI of the users independently of their conscious cultural propositional attitudes. However, this hypothesis has to be verified in future studies. Nevertheless, *some cultural interaction indicators are expected to be valid for HCI design in general* because there are culturally sensitive variables that can be used to measure cultural differences in HCI only

by counting certain interaction events without the necessity of knowing the semantic relations to the application. Such indicators are e. g. number of mouse moves, number of breaks in the mouse movements, speed of mouse movements, number of mouse clicks, number of interaction breaks, and possibly the number of acknowledging or refusing system messages. Surely, all those indicators can also be connected semantically to the use cases or applications. However, the values of the cultural interaction indicators change in a similar way even if different use cases and test tasks are applied. Hence, simply counting such events related to the session duration from users of one culture and comparing them to users of another culture is obviously sufficient to indicate differences in the interaction behaviour of culturally different users with the system. Further research should bring forward more details.

## **5 Conclusion**

The intercultural interaction analysis tool serves to record the user's interaction with the computer to be able to identify cultural variables like color, positioning, information density, interaction speed, interaction patterns, and their values, which enable the deduction of design rules for cross-cultural HCI design. The analysis of the collected data in two online studies with Chinese, German and English speaking employees of SiemensVDO all over the world using this tool showed that there are correlations between the interaction behaviour of the users with the system and their cultural background. The following reciprocal confirming aspects of the two studies quantitatively and qualitatively attest the good reliability and criteria validity of the statistical results received in these two studies:

- High discrimination rate by the cultural interaction indicators of over 80%,
- High accordance of the cultural interaction indicators found by applying different statistical methods,
- High correlated quantitative comparable results of two separated studies.

Moreover, several results presented in this paper are in accordance with other studies, which support their mutual correctness of methodology and outcome (e. g. [Vöhringer-Kuhnt 06], [Kralisch 06], [Kamentz 07]). There are different interaction patterns according to the cultural background of the users ((C) vs. (G) or (E)) regarding e. g. design (complex vs. simple), information density (high vs. low), menu structure (high breadth vs. high depth), personalization (high vs. low), language (symbols vs. characters) and interaction devices (no help vs. help). Furthermore, the results imply that the recognition and classification of cultural interaction

patterns in HCI can also be done quantitatively. This is a precondition for adaptability in the sense of the automatic adaptation of the system to the user by the system itself (through monitoring and evaluating the user interactions to be able to implicate the right adaptation) (cf. [Heimgärtner 05] and [Heimgärtner 06]). Hence, this work contributes a good part to establish cultural adaptability in user interfaces by determining cultural differences in HCI concerning interaction patterns. More detailed studies must show whether changing the metrics of potential indicators (or using them in other situations, use cases or circumstances) will improve their discriminating effect and yield appropriate values accordingly to show the *general validity* (i.e. the independency from applications or use cases) of some cultural interaction indicators. Moreover, future studies have to be done to yield relevant cultural variables according to other user groups (e. g. elderly vs. younger people, experienced vs. beginners, female vs. male, drivers of different vehicles etc.).

## 6 Outlook

The validity of the methods implemented in the IIA tool proved by the results of the two online surveys justifies and encourages optimizing and using this tool for more detailed studies in future to refine and explore new cultural interaction indicators. The near-term objective is to develop enhanced techniques for the IIA data analysis module using statistical methods (factors analysis, structure equation models, cluster analysis etc.), data mining, and semantic processing to extract the cultural variables and its values as well as the guidelines for cross-cultural HMI design in a more automatic way. Moreover, the method to implement new use cases easily into the IIA data collection module will be extended (e. g. by employing authoring tools or HMI description languages). Furthermore, applying questionnaires in conjunction with recording biofeedback signals (heart rate and skin response) will give better-controlled insights into the user preferences. The best discriminating algorithms for cultural adaptability found, using the methods mentioned above, can be implemented in, and tested with the IIA data evaluation module to prove their applicability. Qualitative evaluation studies using intercultural usability tests with users of the respective countries also under mental stress e. g. in realistic driving situations (using the IIA tool in combination with a driving simulator) or in real driving situations (in field studies) must verify the usability of cultural adaptability.



## 7 References

- [De la Cruz et al. 05]. De la Cruz, T.; Mandl, T.; Womser-Hacker, C. Cultural Dependency of Quality Perception and Web Page Evaluation Guidelines: Results from a Survey. In: Day D, del Galdo E, Evers V, editors. *Designing for Global Markets 7: Proceedings of the Seventh International Workshop on Internationalization of Products and Systems (IWIPS 2005)*. Amsterdam, The Netherlands 2005. p.15 – 27.
- [Hall 76]. Hall, E.T. *Beyond Culture*. Garden City, N.Y.: Anchor Press, 1976.
- [Halpin et al. 57]. Halpin, A.W.; Winer, B.J. A factorial study of the leader behavior descriptions. In: Stogdill RM, Coons AE, editors. *Leader behavior: Its description and measurement*. Columbus, OH: Bureau of Business Research, Ohio State University, 1957.
- [Heimgärtner 05]. Heimgärtner, R. Research in Progress: Towards Cross-Cultural Adaptive Human-Machine-Interaction in Automotive Navigation Systems. In: Day D, del Galdo EM, editors. *Proceedings of the Seventh International Workshop on Internationalization of Products and Systems. IWIPS 2005*, The Netherlands, Amsterdam: Grafisch Centrum Amsterdam, 2005. p.97 – 111.
- [Heimgärtner 06]. Heimgärtner, R. Measuring Cultural Differences in Human Computer Interaction as Preparatory Work for Cross-Cultural Adaptability in Navigation Systems. In: *Ueware 2006, VDI-Bericht Nr. 1946*, VDI-Verlag, Düsseldorf (2006) 301 – 314
- [Herczeg 94]. Herczeg, M. *Software-Ergonomie. Grundlagen der Mensch- Computer-Kommunikation*. Bonn: Oldenburg-Verlag, 1994.
- [Hofstede 94]. Hofstede, G. *VSM94: Values Survey Module 1994 Manual*. Tilberg, Netherlands: IRIC, 1994.
- [Hofstede et al. 05]. Hofstede, G.; Hofstede, J.G. *Cultures and Organizations: Software of the Mind*. 2<sup>nd</sup> Edition. New York, USA: McGraw-Hill, 2005.
- [Honold 00]. Honold, P. *Interkulturelles Usability Engineering. Eine Untersuchung zu kulturellen Einflüssen auf die Gestaltung und Nutzung technischer Produkte*. VDI-Verlag, Düsseldorf (2000).
- [Kamentz 07]. Kamentz, E. *Adaptivität von hypermedialen Lernsystemen*. Dissertation 2007.
- [Kralisch 06]. Kralisch, A. *The impact of culture and language on the use of the internet – empirical analyses of behavior and attitudes*. Dissertation 2006.
- [Prendinger et al. 04]. Prendinger, H.; Ishizuka, M. *Life-Like Characters*: Springer, 2004.
- [Röse et al. 01]. Röse, K.; Liu, L.; Zühlke, D. Design Issues in Mainland China: Demands for a Localized Human-Machine-Interaction Design. In: Johannsen G, editor. 8<sup>th</sup> IFAC/IFIPS/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems. Kassel: Preprints, 2001. p.17 – 22.
- [Victor 97]. Victor, D. *International Business Communication*. New York: Prentice Hall, 1997.
- [Vöhringer-Kuhnt 06]. Vöhringer-Kuhnt, T. Asiatische vs. europäische HMI Lösungen von Fahrerinformationssystemen. In: *Ueware 2006, VDI-Bericht Nr. 1946*, VDI-Verlag, Düsseldorf (2006). 279 – 287.



# A Wizard-of-Oz Setting for Multimodal Interaction

## An Approach to User-Based Elicitation of Design Patterns\*

*Andreas Ratzka*

Universität Regensburg  
Institut für Medien-, Informations- und Kulturwissenschaft  
93040 Regensburg  
*Andreas.Ratzka@sprachlit.uni-regensburg.de*

### **Zusammenfassung**

Dieser Beitrag beschreibt ein Wizard-of-Oz-Setting für multimodale Interaktion. Dieser Versuchsaufbau ist ein wichtiger Arbeitsbaustein für die Entwicklung einer Entwurfsmethodik für multimodale Interaktion auf Basis von Wiederverwendung und Nutzerpartizipation. Wiederverwendung von Entwürfen geschieht hierbei mit Hilfe von *Design Pattern Languages*, einer Methode zur Verwaltung von Designinformationen für Entwickler und Usability-Experten. Nutzerpartizipation erfolgt über Wizard-of-Oz-Tests, bei denen das Systemverhalten in frühen Entwicklungsphasen durch menschliche Agenten simuliert wird. Als Testszenario dient die Interaktion mit einem multimodalen E-Mail-System für mobile Geräte und für Desktoprechner. Dabei sollen Hypothesen über multimodale Interaktion erkundet und Design-Pattern-Kandidaten identifiziert werden.

### **Abstract**

This paper describes a Wizard of Oz setting for multimodal interaction. The design of this experimental setup is one working package in our research on a design methodology for multimodal interaction based on design reuse and user participation. Design reuse can be accomplished by design pattern languages, which are a powerful means of information management for both system developers and usability experts. User involvement is performed via Wizard-of-Oz tests, which simulate system behaviour by means of human agents. The test scenario is managing e-mail with multimodal mobile and desktop devices. Our goal is to explore hypotheses on mul-

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 159-170.

timodal interaction and thus identify pattern candidates for the design of multimodal interaction.

## I Introduction

This paper describes the research design for Wizard of Oz simulations of multimodal interactive systems and is situated in the context of methodological research on multimodal interaction design. After a thorough review of literature on model-based user interface design (such as Bürgy 2002, Calvary et al. 2003, da Silva 2001, Souchon et al. 2002, Trøttestad 2002, Wilson et al. 1993) following design aspects for multimodal interaction have been presented in previous work (Ratzka 2006, Ratzka & Wolff 2006):

- Cross-application-aspects
- Workflow and task
- Contextual aspects, i.e. user, situation, environment, device
- Adaptation concept
- Application-specific aspects
- Interaction
- Presentation
- Software Architecture

Our methodology is based on design reuse and user participation. Design reuse can be accomplished with so called *design patterns* and *pattern languages*. User participation and early user involvement is done with Wizard of Oz tests.

The term *design pattern* was firstly used by Alexander et al. (1977) in the domain of architecture and later on introduced into the fields of software engineering (Gamma et al. 1995, Trowbridge & Cunningham 2001) and user interface design (Borchers 2001, van Welie & van der Weer 2003). Design patterns describe solutions to well known design problems in a structured and retrievable way. They relate together a striking name, context and problem descriptions, the description of a rather abstract solution, concrete examples, as well as strengths and pitfalls of this solution.

Patterns constitute an information management tool for system and interaction design, especially when they are grouped together as interlinked pattern language, which can be easily explored by the system designer. Furthermore, by means of pattern languages, common design problems and their solutions can be called by their name and can be discussed more easily by both software and usability experts.

Multimodality has proven to be a successful solution, especially for map based tasks and image editing (Oviatt 1996, Cohen et al. 1997, 2000, Ren et al. 2000, Raisamo & R  ih   2000, Gorniak & Roy 2003, Hiyoshi & Shimazu 1994, Milota 2004). Research on multimodal interaction with automotive applications (Neuss 2001, Salmen 2002, Seiffert 2002, Niedermaier 2003) and personal assistants (Bers et al. 1998, Comeford et al. 2001, Miyazaki 2002) indicate the plausibility of multimodality in these contexts as well.

Some of the high level research results gained in these domains, such as the often cited buzzwords *mutual disambiguation* and *redundant information display*, can be formulated as reusable solutions to recurring design problems and thus be described in a pattern format. In previous work, we have identified a still incomplete collection of patterns and pattern candidates for multimodal interaction from literature review (cf. Ratzka & Wolff 2006). The pattern *human action source distribution* will serve as an illustrative example:

### **1.1 Human Action Source Distribution**

<i>Context</i>	The input of different data types (common in image editing systems, such as selecting a tool from a palette and drawing a figure) sometimes requires a repetitive repositioning of the user's hands between mouse and keyboard or a repetitive repositioning of the mouse cursor.
<i>Problem</i>	Repetitive repositioning is non-productive and slows down interaction.
<i>Solution</i>	Partition alternating subtasks into different action channels. Make use of two-handed (Raisamo & R��ih�� 2000) or speech-enhanced multimodal interaction (Hiyoshi & Shimazu 1994, Milota 2004).

Design patterns relate together problems and *well proven* solutions. In contrast to WIMP<sup>1</sup>-based interaction and web-based applications, multimodality has not yet reached a high degree of dissemination in consumer electronics. Only few design questions can be countered with well proven solutions from research. Other research results hold for specialized contexts only and cannot be generalised. Pattern candidates won from them are proven solutions only in that context. This imposes that user participation is needed to verify whether a certain pattern really fits into the target context, whether a pattern candidate can be generalised to a valid pattern. Furthermore multimodal interaction invites to new design ideas which seem plausible but are not proven empirically. That's why our literature review has to be complemented by user involvement, performed by means of Wizard of Oz tests.

---

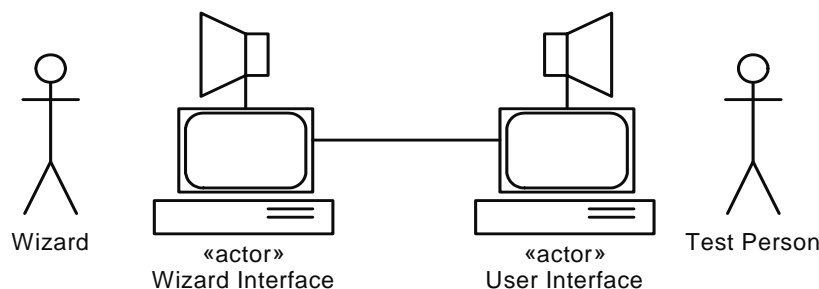
<sup>1</sup> WIMP: **W**indows, **I**cons, **M**enus, **P**ointing device.

## 2 Experimental Setting

Wizard of Oz simulations are a well known approach of informal prototyping in HCI research, especially in spoken interaction and multimodal interaction (cf. Balbo et al. 1993, Salber & Coutaz 1993, Womser-Hacker 1993, Hitzemberger & Womser-Hacker 1995, Cenek et al. 2005, Klemmer et al. 2000). They tend to counter a common problem of human computer interaction researchers when exploring new interface styles: At early stages of interface design, necessary components for interpreting user input are not yet available. The requirements needed to implement them are still unknown, but can be elicited by early user participation accomplished by means of simulations.

Thus Wizard of Oz settings allow user participation at the very first stages of interface design. Alternative approaches of informal prototyping comprise paper mock-ups and storyboarding, which are both exploited in multimodal user interface design (Chandler et al. 2002, Sinha & Landay 2001, Sinha & Landay 2003). In contrast to those techniques, Wizard of Oz tests provide to test users a more realistic look and feel so that design alternatives can be discussed more deeply.

To perform Wizard tests, you need at least two connected computers, one for the user interface, for the test person to perform interactive tasks, and another one for the Wizard, the person who interprets user input and simulates system behaviour. In multimodal scenarios, multiple Wizards may be required, e. g. one for input and one for output or one for each input modality.



*Fig. 1: Wizard of Oz System Setup*

This work focuses on the multimodal enhancement of WIMP-based desktop and mobile interaction with spoken elements. Whereas the WIMP-based behaviour is implemented directly, spoken input and multimodal fusion have to be simulated. The wizard interprets user input and may change the data model of the application accordingly, so that the view of the user interface is updated. He can trigger pre-formulated spoken output, unless an automatic speech feedback strategy is acti-

vated. This requires an efficient wizard interface, which can be controlled in real time via shortcut keys.

The platform simulates a mobile multimodal digital assistant (mobile setting) as well as a multimodal desktop e-mail client (desktop setting). The test person has to perform the task provided by the test scenario, i.e. to organise her incoming mail, organize and delegate meetings etc.

This scenario has been chosen because e-mail is one of the most successful and widespread computer applications (Ducheneaut & Watts 2005, Whittaker et al. 2005). Furthermore, other research (Bers et al. 1998, Comeford et al. 2001, Miyazaki 2002) indicates the plausibility of multimodal interaction for this application domain. In three consecutive surveys (cognitive walkthrough with usability experts, qualitative user study, quantitative user study) hypotheses on multimodal interaction will be examined. In these studies, well determined test scenarios (in the style of *mail baskets*, which are popular exercises in assessment centres) will be combined with free exploration of the user interface by the test person.

One central question to be answered via this method concerns the feedback strategy, especially the interplay and balance between graphical and acoustic system output. Below the two pattern candidates *system initiated modality hint* and *speech-enhanced display* are shortly outlined as examples. The former attempts to counter the conflict between data appropriate modality allocation (such as graphical display of long lists) and interaction history appropriate modality allocation (answering to speech commands via speech replies). The latter one picks up the potential of multimodal interaction to amplify virtually the display of a small device via speech output, so that cluttered screens and navigation complexity can be minimised.

## **2.1 System Initiated Modality Hint**

*Context* Multimodality provides the possibility to select modalities flexibly according to task and data properties, user preferences, and context factors.

*Problem* Sometimes data properties and context factors are in conflict, that means, they encourage the selection of different modalities. Both system and user may lack information for selecting the most feasible interaction style, so that neither pure system initiative nor pure user initiative seem appropriate for modality selection.

*Solution* Point out alternative interaction styles to the user via spoken hints. If the user requests information via speech commands (“get messages from John Long”) and the search result is more easily displayed graphically (as a list

of messages) a spoken modality hint (such as “look at the display or say ‘read out’”) would complement the visual output but preserve interaction flexibility.

## 2.2 Speech-enhanced Display

Context	Some interactive systems support a lot of different tasks. During daily work it is necessary to switch between tasks to get the information needed to perform the main task. When a user wants to answer an e-mail he might need other data which can be found in previous correspondence.
Problem	Small displays make such “multi-tasking” very difficult. Different task windows cannot be displayed all at once, whereas switching between several task screens imposes high navigational efforts, such that users might lose orientation.
Solution	If the display is too small for simultaneous presentation of several task windows, consider partitioning different task displays to different output modalities. Users might listen to new mail or to e-mails of a specific thread while composing a message on the same topic and preserving the visual display.

Both pattern candidates seem plausible but still lack empirical evidence which can be obtained only by user participation. In contrast to full functioning prototypes, Wizard of Oz simulations can react more flexibly to incrementally revealed interaction requirements which are unknown in the very first phases of development.

## 3 Application Architecture

From the perspective of software architecture human computer interaction can be viewed as the creation, selection, manipulation, and destruction of interaction objects. Multimodal interaction imposes the need of specialized interaction objects for covering modality specific interaction possibilities. In order to maintain consistency and to enable synergistic and alternating modality usage, these specialized interaction objects have to be inter-connected through standardized interaction protocols.



That's why the interaction objects of the (simulated) user interfaces are organized in a distributed multi-agent architecture<sup>2</sup> making use of observer, adapter, proxy, and other design patterns. This architecture is inspired by the MVC-extensions of Visual Works Smalltalk (Krasner & Pope 1988, Lewis 1995 p. 103 ff.) on the one hand and PAC-Amadeus (Coutaz 1987, Coutaz 1994, Nigay & Coutaz 1995) on the other hand.

In MVC- and PAC-based architectures observer-objects (here: views) are wrapped around each functional core object (here: model) to get notified about relevant data changes, i.e. to maintain consistency between the functional core and the user interfaces.

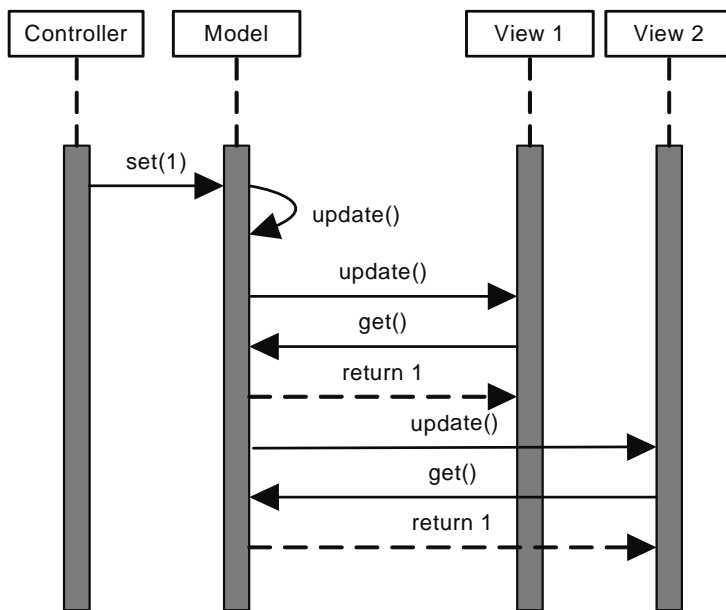


Fig. 2: The Observer Pattern in the MVC Paradigm

Adapters simplify the complex interface of some functional core objects to fit to the MVC protocol.

The proxy-pattern is used in order to introduce one or more degrees of indirection between objects belonging to the functional core of the application and those belonging to the user interface component.

Indirection is required when several interaction objects are combined together, e. g. one for selecting an interaction object and another one for modifying the current selection. This combination of the observer and proxy patterns is also called *subject channel* (view Lewis 1995 p. 103 ff.).

<sup>2</sup> In our case we mean reactive agents and not intelligent agents (cf. Coutaz 1994).

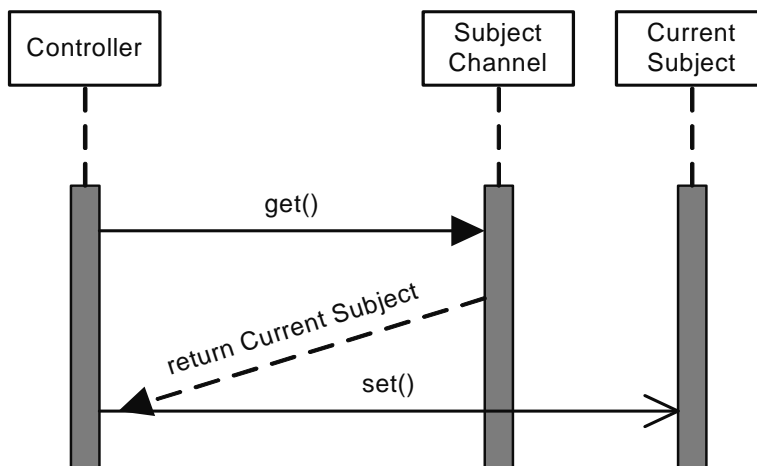


Fig. 3: The Proxy Pattern in extended MVC Architectures with Subject Channels

Further indirection is required to make interaction objects remote-controllable which is needed in a distributed Wizard of Oz scenario but also in real working systems, where the speech recognition component and the user interface are running on separate computers.

## 4 Technological Base

The Wizard of Oz platform is realized as a distributed system consisting of a pocket PC and a desktop PC (mobile setting) or of two desktop PCs (desktop setting) respectively. The communication between the distributed components is realized via socket connections through USB, Bluetooth or Ethernet.

The platform is implemented using *ewe* (cf. Brereton 2005), a Java-based toolkit providing libraries for programming highly portable java applications. The application classes can be run with an ordinary J2SE runtime environment or, and that is necessary for mobile applications, using the Ewe virtual machine.

Applications to be run with the Ewe virtual machine are restricted to the use of the Ewe API (including a subset of java.lang classes), and third party libraries, e. g. your own code.

Applications run with a J2SE runtime environment can combine both standard Java and Ewe libraries. The resulting applications are less portable, but this approach provides more flexibility for implementing the desktop components, i.e. the Wizard interface and the user interface for the desktop setting using JFC (cf. Sun Microsystems 2006), which are easier to handle and more thoroughly documented than the Ewe GUI library.

## 5 Conclusion and Future Work

This paper describes a Wizard of Oz setting in the context of methodological research on multimodal interaction for mobile and desktop devices. Early user involvement performed via wizard tests complements the reuse-based approach of using design patterns. This is important because multimodal interaction is still innovative and therefore lacks a large inventory of proven design solutions. User involvement is needed in design phases as early as possible, which encourages the application of Wizard of Oz simulations.

This research intends to examine design aspects concerning task, context of use, adaptation, interaction and presentation in more detail by means of Wizard of Oz simulations.

In the short run we will perform qualitative user tests in order to validate the general plausibility of design decisions motivated by our pattern candidates. These user tests will be complemented by open interviews and focus group discussions in order to elicit further pattern candidates.

In the long term evaluation tools, eye-tracking as additional interaction modality, as well as a speech engine to allow more scalable user interface simulations have to be integrated in order to allow the collection of reliable quantitative data.

As a by-product, architectural considerations relevant for the design of multimodal user interfaces, which become apparent when implementing Wizard of Oz prototypes, will be collected and further systematised.

## 6 References

- Alexander, C.; Ishikawa, S.; Silverstein, M.; Jacobson, M.; Fiksdahl-King, I.; Angel, S. (1977). *A Pattern Language*. Oxford University Press.
- Balbo, S.; Coutaz, J.; Salber, D. (1993). Towards automatic evaluation of multimodal user interfaces. In: *IUI '93: Proceedings of the 1st international conference on Intelligent user interfaces*. New York, NY, USA: ACM Press, 201-208.
- Bers, J.; Miller, S.; Makhoul, J. (1998). Designing conversational interfaces with multimodal interaction. In: *DARPA Workshop on Broadcast News Understanding Systems*. DARPA, 319-321.
- Borchers, J. O. (2001). A Pattern Approach to Interaction Design. *AI & Society Journal of Human-Centred Systems and Machine Intelligence* 15(4), Springer, 359-376.

- Brereton, M. (2005). *Welcome to ewe. The easiest way to program your mobile PC*. Website: Ewesoft, December 20<sup>th</sup> 2005. Lastly retrieved: April 04, 2007 from: <http://www.ewesoft.com/>.
- Bürge, C. (2002). An Interaction Constraints Model for Mobile and Wearable Computer-Aided Engineering Systems in Industrial Applications. Doctoral Thesis, University of Pittsburgh, Pennsylvania, USA.
- Calvary, G.; Coutaz, J.; Thevenin, D.; Limbourg, Q.; Bouillon, L.; Vanderdonckt, J. (2003). A unifying reference framework for multi-target user interfaces. *Interacting with Computers* 15(3), 289-308.
- Cenek, P.; Melichar, M.; Rajman, M. (2005). A Framework for Rapid Multimodal Application Design. *Proc. of TSD'05*. Karlovy Vary, Czech Republic, Springer, 393-403.
- Chandler, C. D.; Lo, G.; Sinha, A. K. (2002). Multimodal theater: extending low fidelity paper prototyping to multimodal applications. In: *CHI '02: CHI'02 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM Press.
- Cohen, P. R.; Johnston, M.; McGee, D.; Oviatt, S.; Pittman, J.; Smith, I.; Chen, L.; Clow, J. (1997). QuickSet: multimodal interaction for distributed applications. In: *MULTIMEDIA '97: Proceedings of the fifth ACM international conference on Multimedia*. New York, NY, USA: ACM Press, 31-40.
- Cohen, P.; McGee, D.; Clow, J. (2000). The efficiency of multimodal interaction for a map-based task. In: *Proceedings of the sixth conference on Applied natural language processing*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 331-338.
- Comerford, L.; Frank, D.; Gopalakrishnan, P.; Gopinath, R.; Sedivy, J. (2001). The IBM Personal Speech Assistant. In: *Proc. of IEEE ICASSP'01*. DARPA, 319-321.
- Coutaz, J. (1987). PAC, an object-oriented model for dialog design. In: *Proceedings of IFIP INTERACT'87: Human-Computer Interaction*, chapter 2: Design and Evaluation Methods: 2.5 Dialogue Design and Evaluation, 431-436.
- Coutaz, J. (1994). *Software Architecture Modeling For User Interfaces*. In: J. J. Marciniak, (ed.): *Encyclopedia of Software Engineering*. Wiley, Chichester, 38-49.
- Ducheneaut, N.; Watts, L. A. (2005). In Search of Coherence: A Review of E-Mail Research. *Human Computer Interaction* 20, 1-9.
- Gamma, E.; Helm, R.; Johnson, R.; Vlissides, J. (1995). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley.
- Gorniak, P.; Roy, D. (2003). Augmenting user interfaces with adaptive speech commands. In: *ICMI '03: Proceedings of the 5<sup>th</sup> international conference on Multimodal interfaces*. New York, NY, USA: ACM Press, 176-179.
- Grasso, M. A.; Ebert, D. S.; and Finin, T. W. (1998). The integrality of speech in multimodal interfaces. *ACM Transactions on Computer-Human Interaction* 5(4), 303-325.
- Hitzenberger, L.; Womser-Hacker, Ch. (1995). Experimentelle Untersuchungen zu multimodalen natürlichsprachlichen Dialogen in der Mensch-Computer-Interaktion. *SDV – Sprache und Datenverarbeitung*. 19(1), 51-61.

- Hiyoshi, M.; Shimazu, H. (1994). Drawing pictures with natural language and direct manipulation. In: *Proceedings of the 15<sup>th</sup> conference on Computational linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 722-726.
- Klemmer, S. R.; Sinha, A. K.; Chen, J.; Landay, J. A.; Aboobaker, N.; Wang, A. (2000). Suede: a Wizard of Oz prototyping tool for speech user interfaces. In: *UIST '00: Proceedings of the 13<sup>th</sup> annual ACM symposium on User interface software and technology*. New York, NY, USA: ACM Press 1-10
- Krasner, G. E. and Pope, S. T. (1988). A cookbook for using the model-view-controller user interface paradigm in Smalltalk-80. *Journal of Object Oriented Programming* 1(3), 26-49.
- Lewis, S. (1995). *The Art and Science of Smalltalk*. Prentice Hall: London et al.
- Milota, A. D. (2004). Modality fusion for graphic design applications. In: *ICMI '04: Proceedings of the 6<sup>th</sup> international conference on Multimodal interfaces*. New York, NY, USA: ACM Press, 167-174.
- Miyazaki, J. (2002). Discussion Board System with modality variation: From Multimodality to User Freedom. Tampere University, Diploma Thesis.
- Neuss, R. (2001). *Usability Engineering als Ansatz zum Multimodalen Mensch-Maschine-Dialog*. Fakultät für Elektrotechnik und Informationstechnik, Technische Universität München, Doctoral Thesis.
- Niedermaier, F. B. (2003). *Entwicklung und Bewertung eines Rapid-Prototyping Ansatzes zur multimodalen Mensch-Maschine-Interaktion im Kraftfahrzeug*. Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München, Doctoral Thesis.
- Nigay, L.; Coutaz, J. (1995). A generic platform for addressing the multimodal challenge. In: *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems 1: Multimodal Interfaces*. 98-105.
- Oviatt, S. (1996). Multimodal Interfaces for dynamic interactive maps. In: *Human Factors in Computing Systems, CHI '96 Conference Proceedings*. ACM Press, 95-102
- Raisamo, R.; Rähä, K.-J. (2000). Design and evaluation of the alignment stick. In: *Interacting with Computers* 12(5), 483-506.
- Ratzka, A.; Wolff, Ch. (2006). A Pattern-based Methodology for Multimodal Interaction Design In: *Proc. TSD 2006*. Springer, 677-686.
- Ratzka, A. (2006). Combining Modality Theory and Context Models. In: *Proc. PIT 2006*. Springer, 141-151.
- Ren, X.; Zhang, G.; Dai, G. (2000). An Experimental Study of Input Modes for Multimodal Human-Computer Interaction. In: *ICMI '00: Proceedings of the Third International Conference on Advances in Multimodal Interfaces*. London, UK: Springer-Verlag, 49-56.
- Salber, D.; Coutaz, J. (1993). A Wizard of Oz platform for the study of multimodal systems. In: *CHI '93: INTERACT '93 and CHI '93 conference companion on Human factors in computing systems*. New York, NY, USA: ACM Press, 95-96.
- Salmen, A. (2002). *Multimodale Menüausgabe im Fahrzeug*. Lehrstuhl für Informationswissenschaft, Philosophische Fakultät IV, Universität Regensburg, Doctoral Thesis.

- Seifert, K. (2002). Evaluation multimodaler Computer-Systeme in frühen Entwicklungsphasen, Fakultät V – Verkehrs- und Maschinensysteme. Technische Universität Berlin, Doctoral Thesis.
- da Silva, P. P. (2001). User Interface Declarative Models and Development Environments: A Survey. In: P. Palanque and F. Patern (Eds.): *Proc. DSV-IS'00*. LNCS 1946, Springer, 207-226.
- Sinha, A. K.; Landay, J. A. (2001). Visually prototyping perceptual user interfaces through multimodal storyboarding. In: *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces*. New York, NY, USA: ACM Press, 1-4.
- Sinha, A. K.; Landay, J. A. (2003). Capturing user tests in a multimodal, multidevice informal prototyping tool. In: *ICMI '03: Proceedings of the 5<sup>th</sup> international conference on Multimodal interfaces*. New York, NY, USA: ACM Press.
- Souchon, N.; Limbourg, Q.; and Vanderdonckt, J. (2002). Task Modelling in Multiple Contexts of Use. In: P. Forbrig et al. (Eds.): *Proc. DSV-IS'02*. LNCS 2545, Springer, 59-73.
- Sun Microsystems (2006). *The Java<sup>TM</sup> Tutorials. Trail: Creating a GUI with JFC/SWING*. Lastly retrieved: April 04, 2007 from: <http://java.sun.com/docs/books/tutorial/uiswing/>.
- Trøttestad H. (2002). *Model-based User Interface Design*. Doctoral Thesis, Department of Computer and Information Science, Norwegian University of Science and Technology.
- Trowbridge, D.; Cunningham, W.; Evans, M.; Brader, L.; Slater, P. (2004). *Describing the Enterprise Architectural Space*. Microsoft Corporation, June 2004.
- Van Welie, M.; van der Weer, G. C. (2003). Pattern Languages in Interaction Design: Structure and Organization. In: *Proc. Interact'03*.
- Whittaker, S.; Bellotti, V.; Moody, P. (2005). Introduction to This Special Issue on Revisiting and Reinventing E-Mail. *Human Computer Interaction* 20, 1-9.
- Wilson, S.; Johnson P.; Kelly, C.; Cunningham, J.; and Markopoulos P. (1993). Beyond Hacking: a Model Based Approach to User Interface Design. In: *Proc. HCI'93*.
- Wolff, Ch. (2005). Media Design Patterns. In: Eibl, M.; Womser-Hacker, Ch. & Wolff, Ch. (Eds.). *Designing Information Systems*. Constance: UVK, 209-217.
- Womser-Hacker, Ch. (1993). Statistical Experiments on Computer Talk. In: *First International Conference on Quantitative Linguistics*. Dordrecht et al.: Kluwer, 251-263.

# From Personal Information Management Towards Collaborative Information Management in Enterprises

## Analysis, Solution Concept and Description of a Prototype\*

*Monika Maßun*

University of Regensburg  
Institute for Media, Information, and Cultural Studies (IMIK)  
Universitätsstr. 31  
D-93053 Regensburg  
*monika.massun@stud.uni-regensburg.de*

### Abstract

Enterprises as well as employees struggle to encounter solutions to efficiently handle increasing amounts of electronic information. Most information management systems today are based upon top-down, mainly technical approaches and impose their own structures and procedures. This often results in an even greater confusion as complexity grows and integration with existing systems is difficult. Therefore, in this work, a bottom-up conceptual approach centering on users' needs is developed. It will be shown that enabling users to better – i.e. more flexibly – organize their personal information objects<sup>1</sup> consequently leads to a benefit on enterprise level. The international diagnostics company Roche Diagnostics is taken as a representative showcase to analyze the challenges on the way towards efficient information handling and to design a prototype responding to these challenges.

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 171-187.

<sup>1</sup> Personal information objects are e. g. emails, documents, bookmarks, contacts, calendar entries, tasks etc. needed to perform a specific task. The term will be used consequently in this paper to refer mainly to the first three.

## I Introduction

In order to analyze the information landscape and processes at Roche Diagnostics, in 2003 an external consulting company interviewed a small department of 30 employees. Users felt that most problems resulted from email overload, redundant documents in various locations and suboptimal structures in information repositories. The consultants hereupon suggested to implement a collaborative portal technology in order to unify information access<sup>2</sup> and decentralize content management. However, this technology addresses the core problems (cf. chapter 2.3) only partly. The overall tendency to move content from local to online web-based repositories is clearly irreversible. Still, content has to be structured and maintained somehow. As one of the core problems, namely rigid structures is not addressed, portal technologies were not asserted as a comprehensive solution. Additionally, they were seen by users as yet another complex system they would have to learn how to use<sup>3</sup>.

Instead, the solution was sought in enabling users to better and more flexibly organize their information on an individual level, thus influencing the existing structures and applications in a bottom-up manner. Important fields of research to support this approach are Personal Information Management (PIM) and the collaborative paradigms of Web 2.0.

Chapter 2 contains a detailed analysis of the existing problems and their causes. In Chapter 3 relevant research is discussed and evaluated regarding responses offered to the analyzed problems. Lastly, in chapter 4, the solution concept and the resulting prototype will be described.

## 2 Analysis: Challenges in Information Management

Before starting to analyze emerging problems caused by electronic information overload in enterprises, a basic distinction should be drawn between structured and unstructured data (Table 1). Structured data is usually organized systematically in databases and can be retrieved through syntactical metadata. On the other hand, unstructured data – expressing analogous concepts – is stored rather arbitrarily in file structures or content repositories and can be retrieved through semantic meta-

---

<sup>2</sup> Internal document, not meant for publishing.

<sup>3</sup> This has been expressed in most of the interviews conducted for the doctoral thesis and is in accordance with various experts' opinions at Roche Diagnostics' business area Centralized Diagnostics R&D.



data (Gilbert and Friedman 2006). Unstructured data is estimated to constitute 80% of the information in a company (Murphy 2006, Raghavan 2002). This work focuses mainly on the management of unstructured data, especially personal information objects.

	<i>Structured</i>	<i>Unstructured</i>
<i>Representation</i>	Discrete (rows and columns)	Binary large objects: Less-defined boundaries, less-easily addressable
<i>Storage/Persistence</i>	DBMS or file formats	Unmanaged, file structures or content repository
<i>Metadata Focus</i>	Syntax (for example, location and format)	Semantics (descriptive and other markup)
<i>Integration Tools</i>	ETL, federation/EII and replication	Content integration and content adaptors
<i>Standards</i>	Structured Query Language, Open Database Connectivity and Java Database Connectivity	Java Specification Request, and Information and Content Exchange

*Table 1: Distinction of structured and unstructured data.  
(Source: Gilbert and Friedman 2006)*

## **2.1 Case Study: Roche Diagnostics**

For this doctoral thesis, Roche Diagnostics' business area Research and Development (R&D) has been used as a representative example for the emerging problems managing unstructured data. A survey conducted among 15 employees from different departments and functions gave an impression of the users' perspective about dealing with electronic information on a daily basis<sup>4</sup>. Questions were asked about several data repositories as Outlook, personal and departmental file shares, the portal platform, the document management system and the intranet. It could be observed that emails, documents and intranet pages are the information objects people mostly deal with<sup>5</sup>. Searches for information objects are mainly performed with the search engines offered by the different data repositories. As content-based search is not offered by default search engines like Windows search, users' usually search for filenames and in subfolders. Content-based desktop search was not very familiar among the participants. A unified access to information was sought by 6 out of 15 people, especially in the upper hierarchy (project leaders, middle management), by naming folders the same way in different repositories<sup>6</sup>. Furthermore, there were a

<sup>4</sup> The survey was conducted in 2005.

<sup>5</sup> This is also the focus of PIM research.

<sup>6</sup> This corresponds with results of a study conducted by Boardman and Sasse (2004).

couple of terms that were used repeatedly as folder names by various people independently from each other. The main problems felt by the participants were:

- Too many data repositories
- Deep folder hierarchies, forgotten files
- Imposed structures do not meet the needs
- Different views on information cannot be expressed
- Finding latest versions of forms / templates in shared repositories
- Search for synonyms is not supported
- Data redundancy
- Isolated solutions, no common platforms
- Time-consuming search for information objects

## **2.2 Costs of Sub-Optimal Information Handling**

In order to organize and describe the problems caused by inefficient information management the notion of costs will be used in this work. It has been inspired by a discourse of Jones (2004) and further extended for this work. Jones describes the intellectual and monetary costs generated by two scenarios: keeping useless information and not keeping useful information. The emerging costs are e. g. a person's time and effort spent in keeping or finding information, the clutter of useful and useless data and the thereby increased possibility of not retrieving relevant information. Both the scenarios and the types of costs will be complemented by some more aspects for this work. Three categories of increasing costs are distinguished. Those are...

- ...the user's increased time and effort to handle information,
- ...the loss of information work and
- ...the increasing need of storage volume.

Regarding the last point, many people assume that the costs caused by increased storage volume are not significant due to the relatively low prices of storage hardware. What adds up to those costs, however, is the whole process architecture around it, including the administration, archiving, backup and protection of growing data volumes. Other strong arguments to keep only relevant data and thus reduce storage volume are increased retrieval difficulties in vast data amounts and legal directives for archiving periods (Schlüter 2004). Thus it is here considered as a notable cost factor<sup>7</sup>.

---

<sup>7</sup> Although it is seen as a minor factor by Jones (2004).

### 2.3 Causes of Increased Costs

In this work three causes for these costs have been identified: fragmentation, rigid structures and isolation. They mirror both the findings in literature and the users' perspective expressed through the survey results (cf. chapter 2.1). The relation between causes and consequences is not monocausal but can be rather imagined like a matrix (see table 2). The remedies for these identified causes will be proposed in chapter 4 after having discussed the relevant state-of-the-art technologies and concepts (chapter 3).

<i>Consequence</i> <i>Cause</i>	<i>Increased Time Effort,</i> <i>e. g. ↓</i>	<i>Loss of Information</i> <i>Work, e. g. ↓</i>	<i>Increased Storage Vol-</i> <i>ume, e. g. ↓</i>
Fragmentation	search in multiple repositories maintenance and navigation in differing organization structures	lack of overview establishment of folder structures from scratch in each application	storing of email attachments in email client and on file server
Rigid Structures	search for latest versions in various folders increased organization effort navigation in deep hierarchies	search for latest version in various folders loss of multiple aspects of information, only one-dimensional	storing of the same information redundantly in various folders
		→ trade-off between redundancy and loss of multiple aspects of information	
Isolation	the same content is structured by various users relevant information has to be searched independently by various users	relevant information is not shared by all users equally valuable structures are not reusable by others	storing of the same file by many users distribution of attachments via mailing lists

Table 2: Causes and consequences of high costs

#### 2.3.1 Fragmentation

Fragmentation describes the heterogeneity of data formats and applications. Over the last decade, enterprises have implemented numerous data repositories, platforms and applications with partly redundant functions to manage unstructured data (Murphy 2006). Among them are e. g. Enterprise Content Management Systems

(ECM), collaborative/portal platforms, Document Management Systems (DMS), Content Management Systems (CMS), workflow systems and intranets. Only recently companies have started to struggle for a decrease of complexity and the reduction of the number of their applications (Murphy 2006). One consequence of fragmentation is the user's increasing **time effort** to search and organize information which grows in proportion with the number of different storage systems and applications. Users have to manage, maintain and switch between differing structures, search mechanisms and graphical user interfaces. In order to compile a subset of logically linked information objects, generally multiple searches in various systems have to be performed (Karger and Jones 2006).

Also, the fragmentation of information among information silos (Blechar 2005) makes it harder to gain a comprehensive overview over existing and valid documents belonging to one topic. Thus needed information might be retrieved and re-used only with difficulty or not at all which leads to a **loss of information**. Another consequence of the poor overview over existing information is **redundancy** as the same information is stored in several locations by several users in order to assure future retrieval.

### 2.3.2 *Rigid Structures*

Both organizational as well as personal information management systems offer ways to structure content. Usually, these structures are based on tree-like hierarchical folder structures, e. g. Microsoft SharePoint Portal Server, Outlook or file systems. They allow only one parent node (=folder) per child (=information object). In many cases, however, an information object belongs to several topics at a time. The negligence of those manifold aspects of information leads to **information loss**. As a response to the problem of inflexible folder structures, information objects are often duplicated by users in order to have them available in various contexts (=folders) and contribute thereby to increased storage need through **redundant** documents (Schlüter 2004). Eventually, if the same information object is stored in several folders it gets harder to identify the most recent version which contributes again to **increased time effort** for search.

### 2.3.3 *Isolation*

Information can be managed by administrators or by users. Usually, only information managed by administrators is accessible by everyone. A considerable share of information organization, however, takes place on an individual level or in some cases also on group level and is task-based (Boardman and Sasse 2004). Field studies have shown that people invest a lot of time and effort in thinking about how to or-

ganize their information best, especially in their file structures (Boardman and Sasse 2004, Jones et al. 2005). In most cases, these micro-organization efforts remain invisible to the rest of the employees and lead to a **loss of information**. Therefore users have to search independently from each other for information that might have been already retrieved by others. This accounts for an **increased time effort** on an organizational level. File multiplication e. g. due to distribution of information via mailing lists – which is today the most common way to publish information – is one prominent cause for **increased storage volume** (Schlüter 2004).

### 3 State of the Art

Among the research areas answering different aspects of the above discussed problems are e. g. Knowledge Management<sup>8</sup>, Information Retrieval<sup>9</sup>, Web 2.0 and Semantic Web as well as Personal Information Management (PIM). As Web 2.0, PIM and the sub-areas metadata and taxonomies in Information Retrieval have been identified as solution-relevant for this work, they will be discussed in more detail. An overview over their strengths and weaknesses can be found in Table 4 at the end of this chapter.

#### 3.1 Web 2.0

The term Web 2.0 was coined by Tim O'Reilly and people around him in 2004<sup>10</sup>. They used it to describe currently emerging web technologies that focus strongly on user participation. Only two examples of Web 2.0 concepts shall be mentioned here as they are relevant for the solution. These are Wikis and Collaborative Tagging. Wikis allow a group of authors to collaborate on articles. Content creation is itera-

---

<sup>8</sup> Knowledge Management in a broader sense offers on the technological side e. g. Content Management Systems, Document Management Systems, Collaboration / Portal software, Groupware and Workflow systems. Those are seen as further technologies that remain in the paradigm of hierarchical structuring and therefore are not seen as solutions for the described problem focus of this work.

<sup>9</sup> Information Retrieval investigates e. g. Textmining, Visualization, Topic Maps, Automatic Classification, Taxonomies and Metadata. Except the taxonomy and metadata concepts, they have been excluded as solutions as there are already advanced technologies offering this top-down support.

<sup>10</sup> The expression Web 2.0 is quite controversial among experts. Critics like Tim Berners-Lee say that Web 2.0 does not convey anything new but rather reuses ideas that were already part of Web 1.0 (Berners-Lee 2006). However, in this article the term will be used to refer to the typical developments that are usually cited when talking about Web 2.0, mainly participation.

tive as people keep refining and updating the articles. A famous example for a wiki is “Wikipedia”<sup>11</sup>, an encyclopedia established by an unlimited group of volunteers. Despite the liberal collaboration practice the quality of Wikipedia articles is competitive. This was the result of a study by the “Nature” magazine published in 2005 in which experts compared the accuracy of 42 science entries from Wikipedia and the Encyclopedia Britannica (Giles 2005 cited in McAfee 2006). This suggests that some form of self-regulation regarding quality and convergence might emerge in the collaborative process.

“Collaborative Tagging describes the process by which many users add metadata in the form of keywords to shared content.” (McAfee 2006) This method of collaboratively “(re)-indexing” content has recently grown popular on the Web. A known example for collaborative tagging on the Web is the bookmarking service “Delicious”<sup>12</sup>. Organizing content by assigning keywords is not a new idea, however, the collaborative aspect is. Usually, some kind of “authority” assigns keywords or they are derived from the document itself. With tagging any user can assign their individual and free keywords to content. This is especially useful for content that is either too much to be indexed intellectually or when there is nobody in the role of the librarian (McAfee 2006). Organizing one’s content with the help of keywords first of all benefits oneself; the benefit for others is a by-product. The synergy of personal and public benefits may lead to a high motivation to use such a system.

### 3.2 Personal Information Management (PIM)

PIM is a research area that strives to enhance the individual’s possibilities to manage their own electronic information objects. However, enterprises usually place more emphasis on general Knowledge Management and strategic information technologies and PIM is left to the individual to cope with. There are diverging theories: some research seeks the solution in better search engines, others in better structuring possibilities for information or in the combination of both.

#### 3.2.1 Structuring and Unification

“Users need ways to unify, simplify, and consolidate information too often fragmented by location, device, and software application.” For Karger and Jones (2006) information fragmentation represents the main problem when it comes to PIM. Most current PIM systems open up a new world of data formats and structures and thereby contribute to an even higher degree of data fragmentation (Jones 2004,

---

<sup>11</sup> <http://www.wikipedia.org>

<sup>12</sup> <http://www.del.icio.us>

Karger and Jones 2006). The primary goal is therefore to unify personal information and not add to fragmentation. Data can be unified on three levels, the **data format level**, the **display level** and the **semantic level**. Most benefits are won with the semantic level strategy (unified namespace).

<i>Approach</i>	<i>Example</i>	<i>Operations</i>	<i>Enables</i>
standard datatype	text	cut/copy/paste	unified search
unified presentation	window manager, Wincuts	layout, tile, show, hide	simultaneous view of information
unified namespace		reference, dereference	list below
grouping	directories, Taskmaster	group, ungroup	organizing, browsing, simultaneous view
metadata	ID3 tags, file metadata, XML	annotate, query	search, organize, browsing
cross-reference	Web, OLE	link, traverse, embed	simultaneous view, orienteering
relations	RDF, Haystack	record named relationships	unified search, browsing, orienteering

Table 3: Approaches to unification. Source: Karger and Jones 2006.

Karger and Jones (2006) come up with two solution approaches. The first one is the **Universal Labeler (UL)** that allows users to assign a common structure to documents in various applications. It is basically a project plan linked to the original sources. Thereby it combines information management with task or project management. The effort put into project planning by the user can thus be reused directly as information organization structure. The second approach is the **Haystack Project** which offers common nomenclature and graphical user interface to organize various document types independently from the application in which they had been created. It simulates thus a web-like navigation paradigm.

### 3.2.2 Combination of Search and Structuring

The research prototype “Phlat”<sup>13</sup> is an extension for Windows Desktop Search (Cuttrell et al. 2006). It merges desktop search and browsing by enriching information with associative and contextual clues like milestones in time, persons, etc. As a new feature, it offers personal tagging across storage systems.

## 3.3 Metadata Management

One way to achieve data unification through a unified namespace is metadata. Metadata can be considered as an approach supporting multiple classifications. The

<sup>13</sup> <http://research.microsoft.com/adapt/phlat/default.aspx>

Gartner Research Group investigated the significance of metadata for companies. In one of their studies (Blechar 2005) they come to the conclusion that companies need to focus on consistency, reusability and accessibility of metadata in order to improve efficiency regarding time-to-market of new products. Organizational silos, data redundancy and inconsistencies across those silos are the big challenges. A way to overcome those silos is the consistent use of metadata. However, metadata is currently stored in multiple formats and locations. Despite these findings only a minority of the businesses have implemented a metadata management solution, according to Blechar (2005).

Most data integration and federation tools focus on structured data (Gilbert and Friedman 2006). However, there are some developments towards integrating unstructured data as well. IBM and Oracle are working on systems that put a virtual layer on top of heterogeneous repositories and to provide a control point for metadata. Their goal is to deliver a consistent view on information (Murphy 2006). One group of tools are “ETL” tools (Extract, Transform, Load) which support acquisition and integration of multiple data sources and then deliver the data to one or more target databases. One example for this technology is IBM WebSphere Data Stage. Another product group is called “Virtual Data Federation” or “Enterprise Information Integration (EII)” and enables users to integrate data from multiple sources into a single virtual view while the data remains at the source. An example is IBM WebSphere Information Integrator. The newly emerging tools put stronger emphasis on understanding semantics and discovering metadata as well as inference mechanisms. Examples for vendors are Unicorn and Metatomix (Gilbert and Friedman 2006).

### 3.4 Conclusion

Each of these approaches addresses only a subset of the discussed problem causes (fragmentation, rigid structures, isolation; cf. chapter 2.3). The discussed applications are evaluated as follows:

<i>Concept</i>	<i>Example Application</i>	<i>Fragmentation</i>	<i>Rigid Structures</i>	<i>Isolation</i>
Wikis	Wikipedia		(+)	+
Collaborative Tagging	Delicious		+	+
Namespace Unification	Haystack / UL	+	(+)	
Desktop Search + Tags	Phlat	+	+	
Metadata Management	IBM Websphere Data Stage	+	+	

Table 4: Evaluation of state-of-the-art approaches



## 4 Increasing Efficiency of Information Work

As the discussed systems do not meet all identified needs at a time, a new approach has been developed. In this chapter it will be described how the proposed desktop application prototype “TagIt” addresses three identified causes (cf. chapter 2.3). Firstly, the access to and handling of fragmented information objects should be **unified** to enable a consolidated, cross-tool view on personal information in the sense of „one-stop-shopping“. Secondly, **multiple classification** should be possible in order to allow graph-like, flexible and extensible structures. Thirdly, a **centralized, participative approach** should be part of the concept. Instead of keeping the information organization efforts within the reach of a few, it should be possible for everybody to reuse this valuable intellectual work in some way.

In contrast to centralized top-down approaches of Knowledge Management, here the user is in focus. Users are provided in the first place with a tool to more efficiently manage their personal information. As a by-product, this knowledge management activity is available for the rest of the company<sup>14</sup>.

### 4.1 TagIt: From Idea to Implementation

The proposed concept combines several aspects of the discussed state-of-the-art applications. It incorporates cross-tool data unification on a semantic and display level, multiple categorizations through participative keyword assignment and a centralized, reusable repository of metadata. This metadata is constituted by standardized (=terms from existing agreed corporate taxonomies chosen by the user from pick lists) and user-generated (=ad-hoc freely created terms) tags which is considered a very new and innovative characteristic of TagIt. We chose to design a desktop instead of a web-based application because of its more comfortable use and better performance – this is meant to lower the barrier to use it. The application should be easily integrable into daily work processes.

One objective in the project was to reuse the already available tools and resources instead of replacing them. There are already a lot of applications to manage information objects. On the other hand there are existing taxonomies. What is missing is the link between them. The TagIt application provides the possibility to link both corporate and private information objects virtually from existing data repositories

---

<sup>14</sup> As far as corporate content is concerned. Of course, private information will be only retrieved by its owner.

to tags that are either chosen freely or from corporate taxonomies<sup>15</sup>. It can be understood as a “semantic umbrella” (see figure 1).

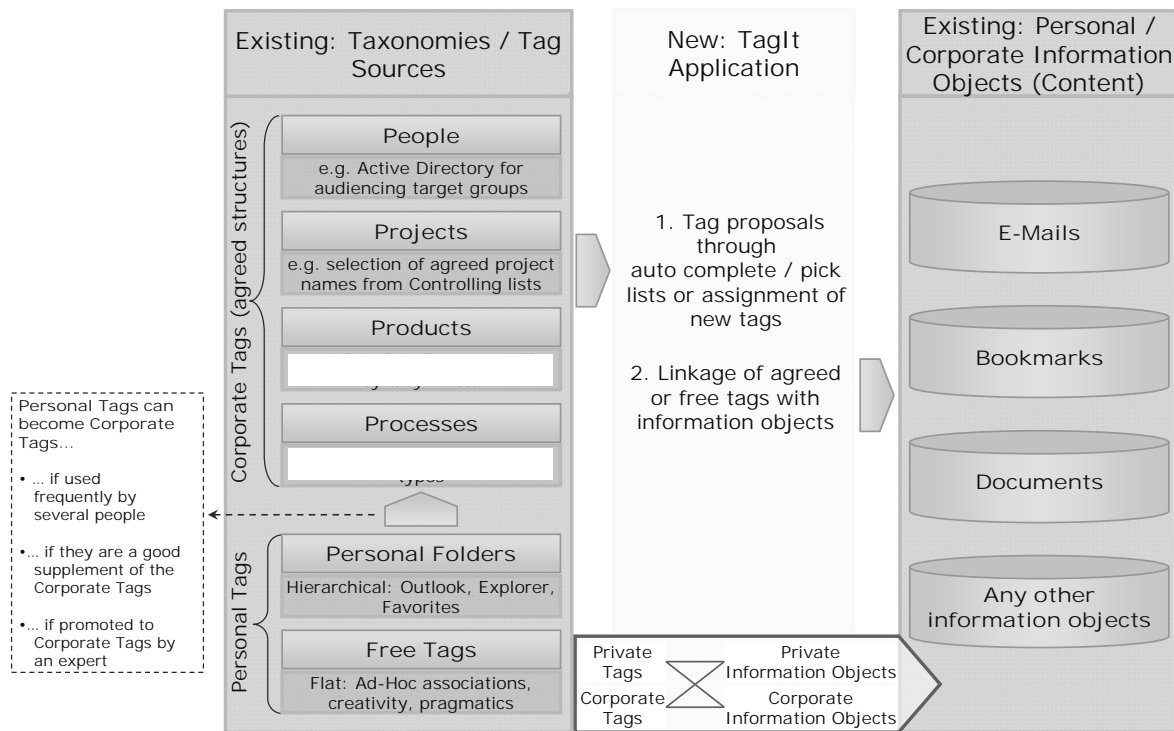


Figure 1: Existing and new components in the Roche environment

#### 4.1.1 Cross-Tool Data Unification Instead of Fragmentation

On the main screen of the TagIt application (see figure 2) there is a three-pane view on the most frequently used information objects (email, documents and Web/intranet pages). This allows users to manage their information objects in one consistent environment instead of switching between email client, file explorer and web browser.

On the left pane there is a navigation tree including corporate and private taxonomies. The number of tagged documents is also displayed. Corporate taxonomies are imported from existing applications, e. g. product names. Private taxonomies are a user's structured tags, i.e. super- and sub-terms can be organized in a hierarchy similar to folder structures. This way of displaying tags was chosen because it is familiar to the user from Windows Explorer and Outlook. The structure can be applied to all information objects independent from their original application. By double-clicking on a tag all associated information objects are displayed in a result list (new tab).

<sup>15</sup> The link between information object and tag is established in a centralized database using the URIs of the information objects.

The upper pane of the screen shows the information object list, either in form of an email, file or web browser. Also the search for tags and the hit list are displayed in this area.

On the bottom pane there is a preview space for information objects. Public and private information objects can be organized in the same virtual repository and are distinguished only by their visibility to other users<sup>16</sup>.

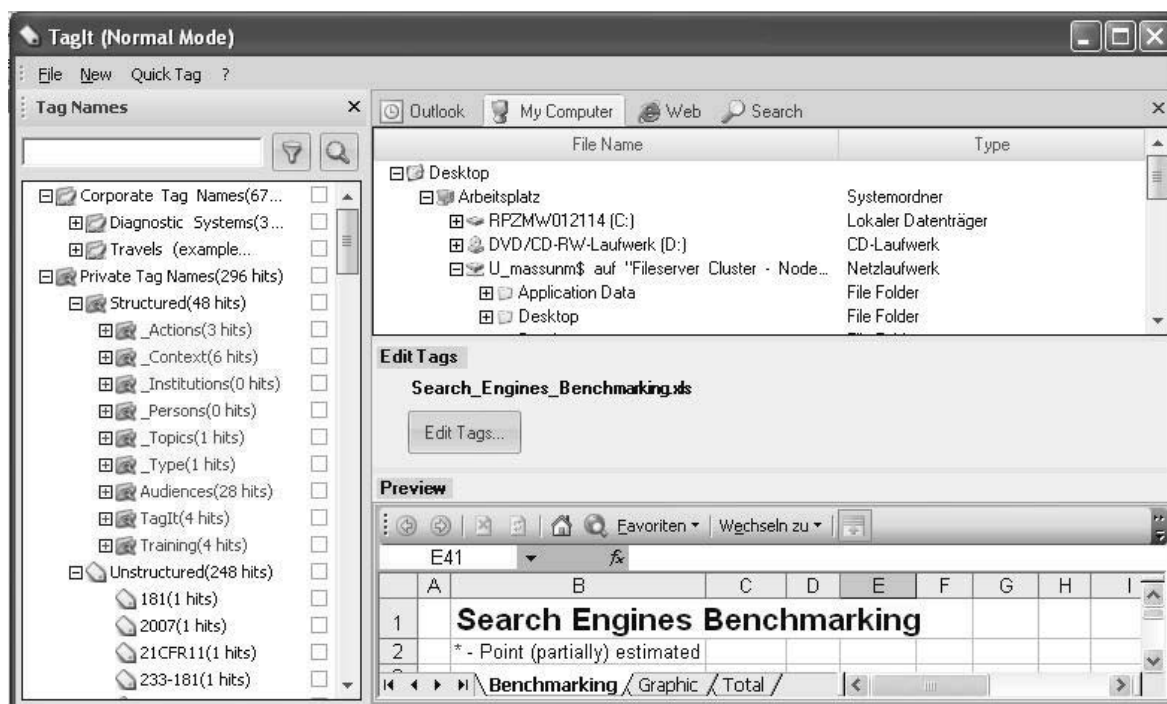


Figure 2: Three-Pane View of the TagIt application

#### 4.1.2 Multiple Classification Instead of Rigid Structuring

As discussed in chapter 2.3.2, conventional folder structures do not allow graph-like structures in which one child can have several parent nodes (=multiple classification). Therefore in TagIt, instead of introducing yet another hierarchical organization structure, the system is extended by the possibility of **tagging**. Tagging allows users to assign multiple keywords (=tags) to information objects. They facilitate a more semantic and better structured storage of information (Lansdale 1988). Private Tags can be flat or hierarchically organized. Although the visual organization of tags can be hierarchical, the links to information objects are multi-hierarchical as one information object can be associated with many virtual folders.

<sup>16</sup> Private information objects like private emails or files on the personal file share are only visible to their owners. Corporate, i.e. public content is visible with all its associated tags to everyone.

Those virtual folders only offer a link to the information object but do not move it physically.

Tags can be used in a consistent way by the help of pick lists or auto-complete functionality in order to support convergence. Furthermore, an easily usable provision of corporate taxonomies is part of the prototype. Users can assign several free or agreed tags to each private or corporate information object and thus locate it in various contexts at a time. Tag assignment is possible either in a dialogue, by dragging & dropping an information object on a tag or with checkboxes (cf. figure 2).

The assignment of tag is also possible from the respective applications (Office applications, Windows Explorer, Internet Explorer) in the form of an Add-In functionality. Clicking a button in these applications invokes a TagIt dialogue and tags can be assigned to the current document.

#### *4.1.3 Centralized User-Generated Metadata Instead of Isolation*

The third conceptual component is the aspect of collective intelligence (Weiss 2005, chapter 3.1). As everybody has his or her own view on information it should be possible for all these views not only to coexist but moreover to contribute to a richer metadata environment accessible by all. This collaborative information network can produce high quality outcomes as shown in chapter 3.1. As mentioned in chapter 2.1 it can be assumed that some kind of convergence emerges as users are already using the same folder names independently from each other.

#### *4.1.4 Search*

Tag-based search is possible in various ways. One can search in private or public information objects or in both. Also, the user can choose whether to search for tags assigned by him- or herself or by other users.

Search possibilities include

- combined searches, i.e. several tags can be connected with Boolean operators,
- search with wildcards,
- search and navigate, i.e. the user can always click on a tag and change the result list by a double click or
- search with suggested terms, i.e. while typing in a search term the autocomplete functionality shows a list of all available tags starting with the typed letters from Corporate and Private Taxonomies as well as other users' tags.

As for security issues, only the information objects a certain user has access to will be retrieved. Private emails or documents of other users are certainly not retrieved.

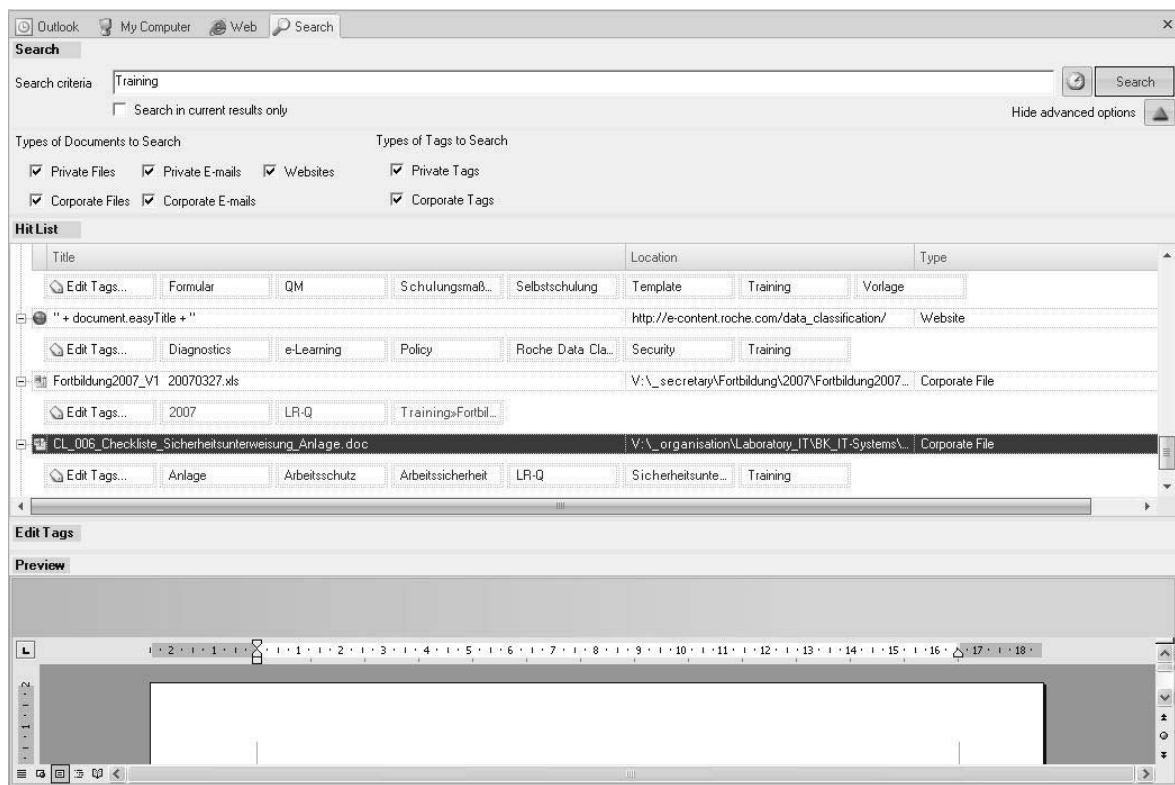


Figure 3: Search interface of TagIt

## 5 Prospects

The pilot will be evaluated in the first months of 2007. This evaluation will be based mainly on user studies. There will be between 10 to 15 pilot users from different areas and in different functions. The target group is project managers and information experts. Ideally, the group of pilot users can be divided into highly structured workers and “chaotic” researchers. Then, the system will be measured against their specific needs. The focus lies on the comparison between effectiveness of information retrieval tasks within the old systems and with the help of the new system. The results may be part of the presentation at ISI 2007.

After the pilot phase the user feedback will be analyzed and used to adapt the system. Apart from this evaluation there are further functionalities that are planned for programming but still have to be analyzed. One important element is the search and tagging with the help of synonym lists provided by a vocabulary server. Also, tags shall be used for ranking mechanisms and the tag-based search shall be integrated with content-based / desktop search.

## 6 References

- Berners-Lee, T. (2006). IBM DeveloperWorks Interview with Scott Laningham. Retrieved July 28, 2006 from <http://www-128.ibm.com/developerworks/podcast/dwi/cm-into82206.txt>.
- Blechar, M. (2005). *How to Manage Your Metadata*. Gartner Research.
- Boardman, R.; Spence, R.; Sasse, M. (2004). *Stuff goes into the computer and doesn't come out: A Cross-tool study of personal information management*. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004) (Vienna, Austria, Apr. 24-29), ACM Press, New York.
- Brynko, B. (2006). *Top 10 Technology Trends*. Retrieved Sept 20, 2006 from <http://pqasb.pqarchiver.com/infotoday/1001218191.html?MAC=a9cc51b746064f49ce95f858997a4fb&did=1001218191&FMT=TG&FMTS=FT:TG:PAGE&date=Mar+2006&author=Barbara+Brynko&pub=Information+Today&printformat=&desc=Top+10+Technology+Trends>.
- Cormier, P. (2006). *Social Tagging For The Enterprise*. Retrieved Sept 20, 2006 from <http://imbok.blogspot.com/2006/02/social-tagging-for-enterprise.html>.
- Cutrell, E.; Robbins, D.; Dumais, S.; Sarin, R. (2006). *Fast, Flexible Filtering with Phlat – Personal Search and Organization Made Easy*. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004) (Montréal, Québec, Canada, Apr. 22-27), ACM Press, New York.
- Dye, J. (2006). *Folksonomy: A Game of High-tech (and High-stakes) Tag*. Retrieved Sept 20, 2006 from [http://pqasb.pqarchiver.com/infotoday/1012273841.html?MAC=4dea07fa3ebce9fd674a357ca2a879c2&did=1012273841&FMT=TG&FMTS=FT:TG:PAGE&date=Apr+2006&author=Jessica+Dye&pub=EContent&printformat=&desc=Folksonomy%3A+A+Game+of+High-tech+\(and+High-stakes\)+Tag](http://pqasb.pqarchiver.com/infotoday/1012273841.html?MAC=4dea07fa3ebce9fd674a357ca2a879c2&did=1012273841&FMT=TG&FMTS=FT:TG:PAGE&date=Apr+2006&author=Jessica+Dye&pub=EContent&printformat=&desc=Folksonomy%3A+A+Game+of+High-tech+(and+High-stakes)+Tag).
- Fichter, D. (2006). *Intranet Applications for Tagging and Folksonomies.*, Retrieved Sept 20, 2006 from <http://pqasb.pqarchiver.com/infotoday/1033181571.html?MAC=7997db9f472d5ea9539fbf409fab7659&did=1033181571&FMT=TG&FMTS=FT:TG:PAGE&date=May%2FJun+2006&author=Darlene+Fichter&pub=Online&printformat=&desc=Intranet+Applications+for+Tagging+and+Folksonomies>.
- Gilbert, M.; Friedman, T. (2006). The New Data Integration Frontier: Unifying Structured and Unstructured Data. Gartner Research.
- Giles J. (2005). *Internet Encyclopedias Go Head To Head*. In: Nature 438, 900-901.
- Golder, S.; Huberman, B. (2006). *Usage Patterns of Collaborative Tagging Systems*. In: Journal of Information Science, 32(2), 198-208.
- Hinchcliffe, D. (2006). *Enable richer business outcomes: Free your intranet with Web 2.0.*, Retrieved Sept 20, 2006 from <http://blogs.zdnet.com/Hinchcliffe/?cat=28>.

- Jones, W. (2004). *Finders, keepers? The present and future perfect in support of personal information management*. First Monday, 9(3) (March 2004). Retrieved Jan 13, 2006 from [http://firstmonday.org/issues/issue9\\_3/jones/index.html](http://firstmonday.org/issues/issue9_3/jones/index.html).
- Jones, W.; Phuwanartnurak, A.; Gill, R.; Bruce, H. (2005). *Don't take my folders away! Organizing personal information to get things done*. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2005) (Portland, OR, Apr. 2-7), ACM Press, New York.
- Karger, D.; Jones W. (2006). *Data Unification in Personal Information Management*. In: Communications of the ACM, 49(1), 77-82.
- Kirk, J. (2006). *Enterprises reap benefits from tagging data*. Retrieved Sept 20, 2006 from <http://www.itworld.com/App/255/060524tagging/>.
- Lansdale, M. (1988). *The psychology of personal information management*. In: Applied Ergonomics, 19(1), 55-66.
- Murphy, B. (2006). *Putting Structure Around Unstructured Data*. Interview with Howard Baldwin. In: Optimize, 51.
- Raghavan, P. (2002). *Information Retrieval for Enterprise Content*. In: Upgrade 3(3), 5-8.
- Schlüter, U. (2004). *Integrationshandbuch Microsoft-Netzwerk*. Galileo Press.
- Weiss, A. (2005). *The Power of Collective Intelligence*. In: netWorker. Beyond file-sharing: collective intelligence, 9(3), 16-23.





# **Competitive Intelligence und Informationsvermittlung in deutschen Unternehmen**

## **Ein Vergleich zwischen Großunternehmen und Mittelstand\***

*Ragna Seidler-de Alwis*

Fachhochschule Köln  
Institut für Informationswissenschaft  
Claudiusstr. 1  
50678 Köln  
*ragna.seidler@fh-koeln.de*

### **Zusammenfassung**

Wettbewerbsdruck und wachsender Informationsbedarf unter anderem auf Grund von Globalisierung und Technisierung führen dazu, dass sich immer mehr Unternehmen mit Themen der strukturierten Informationsversorgung beschäftigen. Der Beitrag zeigt auf, wie unterschiedlich Mittelstand und Großunternehmen ihre Informationsvermittlung und Competitive Intelligence im Unternehmen organisieren.

### **I Warum Informationsvermittlung und Competitive Intelligence in deutschen Unternehmen?**

Jedes Unternehmen sieht sich heute in nahezu allen Märkten mit wachsendem Wettbewerbsdruck und dem damit einhergehenden wachsenden Informationsbedarf konfrontiert. Durch die sich rasch entwickelnde Globalisierung und Technisierung und die damit einhergehende Zunahme der zu bewältigenden Informationen wird wissensbasierte Arbeit immer wichtiger und eine strukturierte Informationsversorgung im Unternehmen dadurch immer notwendiger. Durch den Einsatz von Competitive Intelligence wird das Unternehmen für Veränderungen im Marktumfeld frühzeitig sensibilisiert. Aktivitäten der Wettbewerber können so zum Teil vor-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 189-206.

ausgesagt oder zumindest zeitnah nachvollzogen werden – neue und potentielle Wettbewerber werden rechtzeitig erkannt. Auch mögliche Übernahmeziele können so identifiziert werden. Durch die kontinuierliche Beobachtung ist das Unternehmen in der Lage, aus Fehlern und Erfolgen der Konkurrenz zu lernen. Dabei werden eigene strategische Ziele, Produkte und Prozesse weiterentwickelt und in Bezug auf die Veränderungen des politischen und rechtlichen Umfeldes geprüft und überwacht.<sup>1</sup>

Kenntnisse über die Situation, Pläne oder Strategien der Wettbewerber führen zu einer Optimierung der unternehmerischen Entscheidungen – dies haben sowohl Großunternehmen als auch der Mittelstand in Deutschland erkannt. Der Umgang von CI und der damit zusammenhängenden Informationsvermittlung im Unternehmen divergiert jedoch stark. Das Management muss nicht nur vom Einsatz von Competitive Intelligence überzeugt sein, sondern auch die nötigen finanziellen Mittel und Personalressourcen zur Verfügung stellen. Dadurch besteht vor allem für kleinere Unternehmen aufgrund der langfristig hohen Fixkosten eine Hemmschwelle. Einige Unternehmen setzen Competitive Intelligence noch immer mit Industriespionage gleich und wollen sich deshalb von diesem Instrument distanzieren.<sup>2</sup>

## **2 Competitive Intelligence – Definition des Begriffs**

Competitive Intelligence bezeichnet ein Management-Instrument zur Beobachtung und Analyse des Wettbewerberumfeldes eines Unternehmens. Für Competitive Intelligence werden im deutschsprachigen Raum häufig Begrifflichkeiten wie „Markt- und Wettbewerbsbeobachtung“ oder „Konkurrenzanalyse“ verwendet. „Als ‚Competitive Intelligence‘ (CI) wird einerseits der systematische Prozess der Informationserhebung und -analyse bezeichnet, durch den aus fragmentierten Informationen über Märkte, Wettbewerber und Technologien Entscheidern ein plastisches Verständnis über sein Unternehmensumfeld entsteht. CI-Themen sind dabei meist zukunftsorientierte Aussagen zu Wettbewerberpositionierungen, -intentionen und -strategien. Andererseits ist ‚Intelligence‘ das Endresultat des Prozesses: das benötigte Wissen über Markt und Wettbewerb. Es werden insbesondere Aussagen über die erwarteten Auswirkungen für das eigene Unternehmen und darauf basierende

---

<sup>1</sup> Vgl. Brellochs: Competitive Intelligence u. Knowledge Management, 2000, S.39.

<sup>2</sup> Vgl. Pfaff/Altensen: Competitive Intelligence – Nur eine Domäne der Amerikaner?, 2003

Handlungsempfehlungen getroffen.“<sup>3</sup> Competitive Intelligence liefert ein organisiertes und strukturiertes System zur Analyse der häufig bereits auch im Unternehmen vorliegenden Informationen über die bestehende Konkurrenz.<sup>4</sup> Werden auch interne Daten ausgewertet spricht man nach der Definition der Society of Competitive Intelligence Professionals von Business Intelligence.<sup>5</sup>

### **3 Der deutsche Mittelstand versus deutsche Großunternehmen – eine Abgrenzung**

Die Bedeutung des Mittelstands in Deutschland wird durch Kennzahlen, die vom Institut für Mittelstandsforschung (IfM) Bonn erhoben wurden, eindrucksvoll dargestellt. Demnach umfasst der Mittelstand in der Bundesrepublik Deutschland rund 99,7% aller umsatzsteuerpflichtigen Unternehmen, in denen knapp 70% aller Beschäftigten rund 43% aller Umsätze erwirtschaften und knapp 83% aller Auszubildenden beschäftigt werden.<sup>6</sup> Trotz dieser immensen Bedeutung gibt es für den Mittelstand auch KMU genannt- keine gesetzlich verbindliche oder allgemein gültige Definition. Zur Abgrenzung des Mittelstands werden häufig Kriterien wie die Beschäftigtenzahl, der Jahresumsatz, die Bilanzsumme, die Bruttowertschöpfung oder die Größe der verbundenen Unternehmen herangezogen. In der Praxis ist auch eine Kombination aus mehreren der genannten Kriterien gebräuchlich. Nach §267 HGB „Umschreibung der Größenklassen“ werden kleine und mittelgroße Gesellschaften hinsichtlich der Beschäftigtenzahl, der Bilanzsumme und der Umsatzerlöse abgegrenzt.

Das Institut für Mittelstandsforschung in Bonn hat eine weitere Definition, auf die sich viele Statistiken über den Mittelstand beziehen und die weit verbreitet ist. Hier heißt es: Zum Mittelstand zählen Unternehmen mit einer Beschäftigtenzahl von 10-499 Mitarbeitern, außerdem ist der Jahresumsatz nicht höher als 50 Millionen Euro.<sup>7</sup> Aber auch qualitative Kriterien spielen eine Rolle. Ein wichtiger qualitativer

---

<sup>3</sup> DCIF: Was ist Competitive Intelligence?, 2005.

<sup>4</sup> Vgl. Korn-Ferry / Internat. Univ. of Southern California: Strategies for the Knowledge Economy, Los Angeles, USA 2000.

<sup>5</sup> Vgl. Leitl: Competitive Intelligence, 2005.

<sup>6</sup> Vgl. Inst. f. Mittelstandsforschung Bonn: Mittelstand – 2005.

<sup>7</sup> Vgl. Kayser, G. / Institut für Mittelstandsforschung: Was sind eigentlich kleine und mittlere Unternehmen, 2003.

Aspekt bei der Betrachtung des Mittelstands die enge Verbindung zwischen Unternehmen und Inhabern. So sind KMUs mit aller Regelmäßigkeit Familienbetriebe.<sup>8</sup> Es gibt verschiedene Institutionen, die versuchen Großunternehmen einzugrenzen, wie z. B. das Institut für Mittelstandsforschung, Hoppenstedt oder das statistische Bundesamt. In dieser Arbeit werden Unternehmen ab einem Umsatz von 100 Mio. € als Großunternehmen gewertet, die genannten Quellen dienen als Anhaltspunkte.

Die Unterscheidung von Unternehmen in Kleinstbetriebe, KMU und Großunternehmen erfolgt auf der Basis von Bilanzzahlen (Umsatz) und der Anzahl der Mitarbeiter. Hier wird von einer Mitarbeitergröße von mindestens 500 Angestellten und einem Umsatz von über 50 Millionen Euro gesprochen, die für Großunternehmen charakteristisch sind.<sup>9</sup>

<i>Unternehmensgröße</i>	<i>Zahl der Beschäftigten</i>	<i>Umsatz in € / Jahr</i>
klein	bis 9	Bis unter 1 Million
mittel	10 bis 499	1 bis 50 Millionen
groß	über 500	über 50 Millionen

*Tabelle 1: Übersicht der Unternehmensgrößen, Stand 01.01.2002, Quelle: Institut für Mittelstandsforschung, Bonn*

Qualitative Kriterien sind „divergierende Verhältnisse im Kapitaleigentum“<sup>10</sup>, womit die Institution des Verwaltungsrates auf der einen Seite und dem Geschäftsführer bzw. Vorstand auf der anderen Seite gemeint ist. Dabei obliegt dem Geschäftsführer oder auch CEO bzw. dem Vorstand die operative Führung, der Verwaltungsrat besitzt eine strategische und kontrollierende Rolle.<sup>11</sup> Insgesamt liegt der Anteil von Großunternehmen an allen Unternehmen in Deutschland bei unter 1%. Dennoch beschäftigen die Großunternehmen 41% der Sozialversicherungspflichtigen Arbeitnehmer. Diese Zahlen beruhen auf der Definition von Großunternehmen durch Hoppenstedt. In dieser Definition sind die quantitativen Merkmale eher niedrig angesetzt, Hoppenstedt geht von einem Umsatz ab 20 Mio. Euro aus und identifiziert somit 24.650 Großunternehmen.<sup>12</sup> Laut statistischem Bundes-

<sup>8</sup> Vgl. Bamberger/Evers: Ursachen und Verläufe von Internationalisierungsentscheidungen mittelständischer Unternehmen, 2002.

<sup>9</sup> Vgl. Institut für Mittelstandsforschung, Bonn, 2002.

<sup>10</sup> Vgl. Wikipedia: Großunternehmen.

<sup>11</sup> Vgl. Statistisches Bundesamt, Wiesbaden, [www.destatis.de](http://www.destatis.de)

<sup>12</sup> Vgl. Hoppenstedt: Pressemitteilung zu Großunternehmen vom 20.4.06.

amt existierten in Deutschland 2004 rund 2,96 Mio. umsatzsteuerpflichtige Unternehmen, nur 8.643 erzielten 2003 einen Jahresumsatz größer 50 Mio. Euro und konnten somit als Großunternehmen bezeichnet werden. Größter Wirtschaftszweig in dieser Gruppe ist das „verarbeitende Gewerbe“ mit einem Anteil von 41 % gefolgt vom Bereich „Handel“ mit 32 %. Bezogen auf die 2,96 Mio. Unternehmen ergibt sich für die Großunternehmen ein Anteil kleiner 0,003 %. Die zweite quantitative Größe, die Anzahl der Mitarbeiter deckt sich mit den Umsatzstrukturen. Hier weist das statistische Bundesamt 2003 10.660 Unternehmen aus, die 250 oder mehr sozialversicherungspflichtige Mitarbeiter beschäftigten. Dies entspricht 0,004 % aller Unternehmen. Entsprechend des Statistischen Bundesamtes, gehören 54 % der Großunternehmen als Rechtsform zu den Kapitalgesellschaften, firmieren also z. B. als GmbH oder AG.<sup>13</sup>

### **3.1 Bekanntheit und Akzeptanz von Competitive Intelligence in deutschen Unternehmen**

Der Stellenwert von CI ist in Deutschland vergleichsweise zu den USA geringer. Doch insbesondere in Großunternehmen etabliert sich CI als erfolgreiches Managementtool, im deutschen Mittelstand ist CI jedoch bisher relativ wenig verbreitet. Pfaff und Altensen stellen in Rainer Michaelis Buch dar, dass von den 499 befragten Unternehmen nur 61% der dort befragten Unternehmen den Begriff Competitive Intelligence überhaupt kennen. Nur 32% können die Begrifflichkeit gar korrekt einordnen. Von diesen Befragten waren 48% Mitglieder der Society of Competitive Intelligence Professionals (SCIP). „Ausgenommen der SCIP-Mitglieder, war Competitive Intelligence nur jedem dritten Probanden bekannt“.<sup>14</sup> Bei einer Akzeptanzbefragung stimmten Mitglieder von SCIP relativ deutlich befürwortenden Aussagen zu CI zu – im Gegensatz zu Nicht-SCIP-Mitgliedern, die sich der besonderen Befürwortung enthielten. Ein wesentliches Ergebnis der Studie von Altensen und Pfaff ist, dass CI für die deutschen Betriebe ein nützliches Instrumentarium für die Entwicklung einer Unternehmensstrategie darstellt. 70% der Befragten messen CI und der Strategieentwicklung einen sehr hohen bis hohen Stellenwert bei.<sup>15</sup> Auch die GIA-Studie kommt zu einem ähnlichen Ergebnis. Hier nutzen 83% der befrag-

---

<sup>13</sup> Vgl. Statistisches Bundesamt, Wiesbaden, [www.destatis.de](http://www.destatis.de)

<sup>14</sup> Altensen, A. / Pfaff, D.: Ergebnisse einer bundesweiten Studie. In: Michaeli, R.: 2006, S. 60, 64-66.

<sup>15</sup> Altensen, A. / Pfaff, D.: Ergebnisse einer bundesweiten Studie. In: Michaeli, R.: 2006, S. 67-69.

ten 287 Unternehmen CI sehr intensiv zur Unterstützung der Unternehmensentwicklung und 81% intensiv für den Bereich Marketing und Vertrieb.<sup>16</sup>

Zwischen der Akzeptanz von Competitive Intelligence und der Bereitstellung und von CI-Lösungen, und deren tatsächlicher Nutzung besteht teilweise eine Diskrepanz. Insbesondere in den Bereichen Unternehmensführung, Unternehmensplanung und Vertrieb besteht noch weiteres Nutzungspotenzial.<sup>17</sup> CI-Tätigkeiten werden „in deutschen Unternehmen weitestgehend unzureichend und unstrukturiert ausgeübt: Eine Analyse der Wettbewerber wird nicht kontinuierlich, sondern nur in unregelmäßigen Abständen durchgeführt“.<sup>18</sup> Dies führt unter anderem zu unvollständigen Informationsbeständen. Mangelnde Kenntnisse bzgl. CI-Aktivitäten tragen auch dazu bei. Trotz der Erkenntnis, dass CI einen wichtigen Garant für den Unternehmenserfolg darstellt, werden Aktivitäten „vorwiegend von den Abteilungen Marktforschung und Marketing übernommen – wirklich effizient kann CI hingegen nur durch die Integration einer eigenen Stabsstelle betrieben werden.“<sup>19</sup> Auch Lux und Nösner kommen zu dieser Erkenntnis. Ihrer Meinung nach forciert eine fehlende breite Institutionalisierung im Unternehmen die angesprochene Diskrepanz.<sup>20</sup> Hinzu kommt, dass zwar die meisten Unternehmen den Begriff Competitive Intelligence kennen, diesen aber nicht in den korrekten Kontext einordnen können. Zu diesem Ergebnis kommt eine Umfrage, in der 499 mittelständische und große Unternehmen befragt wurden. Nur 26% können CI korrekt dem allgemeinen Verständnis zuordnen.<sup>21</sup> Das Missverhältnis zwischen Bekanntheit (Akzeptanz) und Nutzung hängt in der Regel auch mit der unternehmensinternen Bedeutung von CI zusammen, die auch mit der Größe der Organisation zusammenhängt.<sup>22</sup> Unternehmen, die mehr als eine Milliarde Euro umsetzen, arbeiten beispielsweise häufiger mit Tools wie Online-Datenbanken und Webrecherchen als andere Organisationen.<sup>23</sup>

---

<sup>16</sup> GIA, White Paper, 2005, S. 20.

<sup>17</sup> Seufert / Lehmann: Business Intelligence – Status quo und zukünftige Entwicklungen 2005, S. 25.

<sup>18</sup> Pfaff / Glasbrenner: Nachholbedarf in Sachen CI, 2004, S. 43.

<sup>19</sup> Pfaff / Glasbrenner: Nachholbedarf in Sachen Competitive Intelligence, S. 43.

<sup>20</sup> Vgl. Lux / Nösner: Competitive Intelligence erfolgreich anwenden, S. 19.

<sup>21</sup> Vgl. Pfaff / Altensen: Nur eine Domäne der Amerikaner? 2003 In: Absatzwirtschaft online.

<sup>22</sup> Seufert / Lehmann: Business Intelligence – Status quo und zukünftige Entwicklungen, 2005, S. 28.

<sup>23</sup> Vgl. Pfaff / Altensen: Nur eine Domäne der Amerikaner? 2003 In: Absatzwirtschaft online.

### **3.2 Einsatz von CI in deutschen Großunternehmen**

Die Verbreitung von CI in Großunternehmen war Untersuchungsgegenstand in verschiedenen Studien der jüngsten Vergangenheit. Diese Studien resümieren einheitlich: CI gewinnt stärker an Bedeutung und findet vor allem Einsatz in großen Unternehmen. Dazu gehört auch, dass diese Unternehmen einen CI-Etat bereitstellen. Trotz der Aufgeschlossenheit gegenüber CI in Großunternehmen, gibt es hinsichtlich der Durchführbarkeit teilweise Zweifel.<sup>24</sup> Und obwohl Markt- und Wettbewerbsanalysen im Rahmen der strategischen Planung und zur frühzeitigen Erkennung von Gefahren im Markt zur Tagesordnung gehören, existiert CI nur in wenigen deutschen Unternehmen als stetiger und strukturierter Prozess mit strategischer Ausrichtung oder in Form einer eigenständigen Funktion.<sup>25</sup>

Im Rahmen einer Studie führte die Fachhochschule Gießen-Friedberg 2003 eine Befragung von 130 überwiegend großen Unternehmen aus dem deutschsprachigen Raum durch. Die Auswertung der Befragung führte zu dem Ergebnis, dass 17% der befragten Unternehmen professionelles CI betreiben. Eine selbständige „CI-Abteilung“ unterhielten knapp ein Fünftel der Unternehmen. Als weitere CI-verantwortliche Abteilungen wurden überwiegend die Bereiche Marketing und Marktforschung genannt, aber auch das Top-Management, F & E und Strategie und Planung wurden stark in entsprechende Aktivitäten involviert. Die Studie unterschied bei der Auswertung der Befragung nach einer Mitgliedschaft und Nicht-Mitgliedschaft bei der Society of CI Professionals (SCIP). Auffällig ist, dass vornehmlich Nicht-SCIP-Mitglieder keine eigenständigen CI-Abteilungen unterhalten. So übernahmen besonders bei Nicht-SCIP-Mitgliedern Abteilungen wie Vertrieb, Einkauf und Finanzen/ Controlling zusätzlich Aufgaben aus dem Bereich der CI.<sup>26</sup>

Eine eigene Befragung des Institutes für Informationswissenschaft der Fachhochschule Köln von 20 Großunternehmen durch Telefoninterviews und einer Online Befragung bestätigt diese Ergebnisse und zeigte zudem, dass ein Großteil der befragten Unternehmen CI-Abteilungen in Form von Non-Profitcentern als Stabstelle eingebunden haben. Lediglich in fünf von den insgesamt 20 auskunftsbereiten Unternehmen ist die CI-Abteilung ein Profitcenter, in einem Unternehmen gab es

---

<sup>24</sup> Altensen, A. / Pfaff, D.: Ergebnisse einer bundesweiten Studie. In: Michaeli, R: 2006, S. 79-80.

<sup>25</sup> Lux, Christian und Nösner, Axel, Wissensmanagement 2004, H. 5, S. 19.

<sup>26</sup> Pfaff, Dietmar: Competitive Intelligence in der Praxis 2005, S.166-167.

laut Auskunft mehrere CI-Abteilungen, die auf unterschiedlichen Ebenen und Bereichen eingebunden sind.<sup>27</sup>

Für Großunternehmen mit einem Nettoumsatz von mehr als einer Milliarde Euro betrug das CI-Budget im Durchschnitt 300.000 €. <sup>28</sup>Auf ein ähnliches Ergebnis kommt die im Rahmen einer globalen Vergleichsstudie der Global Intelligence Alliance (GIA) von Gayer Consulting durchgeführte Befragung unter 25 deutschen Großunternehmen. Obwohl 80% der Befragten ein festes Budget für CI-Aktivitäten zur Verfügung stellten, gaben nur 31% der Unternehmen an, CI in einer eigenständigen Abteilung zu betreiben. Somit hat ein Großteil der befragten Großunternehmen ihre CI-Aktivitäten als Teil einer anderen Funktion, als Einheit einer Stabstelle oder anders organisiert.<sup>29</sup> Eine explizite CI-Funktion besteht seit zehn Jahren und länger bei nur einem Viertel der befragten Unternehmen. 75% haben ihren CI-Bereich erst in den letzten zehn Jahren aufgebaut, das heißt das deutsche Großunternehmen CI als eigenen Funktionsbereich etablieren ist eine erst in den letzten Jahren stetig gestiegene Entwicklung.<sup>30</sup>

Wie auch in der Studie der FH Gießen-Friedberg gibt es auch in der Studie der Global Intelligence Alliance stark divergierende Angaben für die mit CI betrauten Beschäftigten. Es konnte jedoch festgestellt werden, dass mit steigender Anzahl der Mitarbeiter in einem Unternehmen der Anteil der „CI-Vollzeitkräfte“ stark zunimmt.<sup>31</sup> Manfred Bruhn bringt den Erfolg mit der internen Überzeugungsarbeit in Verbindung. Um Mitarbeiter für Dienstleistungen und Methoden zu gewinnen, bedarf „es einer klaren Vision“, die dem gesamten Personal von gut geschulten Mitarbeitern vermittelt werden muss.<sup>32</sup> So muss die Vermittlung der Ziele von CI also ein grundlegender Aspekt des internen Marketings sein. Als schnelle und qualitative Marketinginstrumente fungieren neben unternehmensinternem Fernsehen auch Schwarze Bretter, Mitarbeiterzeitungen und -broschüren sowie Intranet-Lösungen.<sup>33</sup> Der GIA-Studie zufolge wird in deutschen Großunternehmen im Gegensatz zum Großteil anderer europäischer und ausländischer Unternehmen fast

---

<sup>27</sup> Request 2006 – das interne Marketing von CI und InfoDL in Großunternehmen 2006, S. 15.

<sup>28</sup> Altensen, A. / Pfaff, D.: Ergebnisse einer bundesweiten Studie. In: Michaeli, R: 2006, S. 71-72.

<sup>29</sup> GIA, White Paper, CI in large companies 2005, S. 18-21.

<sup>30</sup> GIA, White Paper, CI in large companies, 2005, S. 10.

<sup>31</sup> GIA, White Paper, CI in large companies, 2005, S. 23.

<sup>32</sup> Bruhn, Manfred: Internes Marketing 2. Auflage 1999, S. 74.

<sup>33</sup> Bruhn, Manfred: Internes Marketing 2. Auflage 1999, S. 367-368.



durchweg auf Vollzeitkräfte gesetzt. Allerdings wird dabei weniger Personal verwendet als beispielsweise im asiatisch/pazifischen Raum oder in Nordamerika.<sup>34</sup> Trotz vergleichsweise professioneller personeller Ausstattung und fester Budgetierung befindet sich die CI in Deutschland in seiner Entwicklung gegenüber den USA fünf bis zehn Jahre zurück.<sup>35</sup>

### *3.2.1 Informationsbeschaffung und Informationsvermittlung in deutschen Großunternehmen*

Als Informationsquelle für CI-Abteilungen dienen deutschen Großunternehmen häufig die eigenen Mitarbeiter. Dies gilt beispielsweise für Außendienstler, welche relevante Informationen zum Umfeld des Unternehmens besitzen. 42% der deutschen Konzerne sind dabei der Meinung, dass die Einbeziehung des internen Wissens „gut“ funktioniert. Nur 8% beurteilen diese als „schlecht“. Die Aufbereitung der unternehmensinternen Informationen erfolgt bei 40% der Unternehmen durch interne Datenbanken und Mitarbeiterbefragungen. Auch Newsletter und informelle firmeninterne Netzwerke dienen als Informationsquelle für CI, ebenso Fort- und Weiterbildungsmaßnahmen.<sup>36</sup> Daneben spielt das Data Warehouse bei der Informationsbeschaffung von CI in deutschen Unternehmen eine große Rolle. Nur knapp 4,7% der Unternehmen, die CI-Lösungen in Unternehmen verwenden, wollen langfristig auf ein Data Warehouse verzichten.<sup>37</sup> Newsletter und Mitarbeiterbefragungen dienen Unternehmen allerdings nicht nur zur Gewinnung von Informationen zur Wettbewerbsbeobachtung, sondern auch zur Informationsvermittlung und Bekanntmachung von CI.<sup>38</sup> Eine der wichtigsten Handlungen die Belegschaft über aktuelle Entwicklungen, Veränderungen und Potenziale zu informieren ist auch im digitalen Zeitalter weiterhin der persönliche Kontakt.<sup>39</sup> Genutzt werden die beschafften Informationen bei 54% der untersuchten Unternehmen vorzugsweise für die strategische Ausrichtung und Planung der jeweiligen Organisationen. In Deutschland liegt die Nutzung von Competitive Intelligence für diesen Zweck bei überdurchschnittlich 83%.<sup>40</sup> Wichtige Aspekte sind die Konkurrenz- und Markt-

---

<sup>34</sup> Global Intelligence Alliance White Paper 4/2005, S. 26.

<sup>35</sup> Stippel, Peter: Fast wie beim CIA – Konkurrenzabwehr im globalen Wettbewerb, 2002, S. 18-20.

<sup>36</sup> GIA, White Paper, CI in large companies, 2005, S. 14-15.

<sup>37</sup> Seufert / Lehmann: Business Intelligence – Status quo und zukünftige Entwicklungen 2005, S. 23.

<sup>38</sup> GIA, White Paper, CI in large companies, 2005, S. 11.

<sup>39</sup> Reichwald, R.: Führung, Organisation und Kommunikation 2005, S.132.

<sup>40</sup> GIA, White Paper, CI in large companies, 2005, S. 22.

beobachtung sowie die Sammlung von Informationen über und zu den eigenen Kunden.<sup>41</sup> Besonders intensiv werden CI-Ergebnisse durch das Topmanagement genutzt. Weitere Adressaten für CI sind das mittlere Management und Fachleute – allerdings in einem deutlich geringeren Umfang.<sup>42</sup> Wissensmanagementsysteme werden von Mitarbeitern im Großunternehmen akzeptiert, wenn diese sich im Alltag bewähren. Nach dem Prinzip „Wer will, dass ihm die anderen sagen, was sie wissen, der muss ihnen sagen, was er selbst weiß.“<sup>43</sup> Die Firma Detecon International – ein praktisches Beispiel – baut bei der Einführung von Wissensmanagementsystemen auf neue Regeln der Kommunikation. Den Mitarbeitern werden Studien zugänglich gemacht, die diese zur Evaluierung und Aufarbeitung ihres Wissens nutzen können.

Grundsätzlich bleibt, dass Informationsbeschaffung und Informationsvermittlung von Competitive Intelligence in Großunternehmen stark von ihren Interessensgruppen abhängig ist und dass persönliche Kontakte immer noch eine wichtige Rolle bei der Darstellung eigener Leistungen spielen. Die Überzeugungsarbeit, durchgehend hohe Qualität und Professionalität sind nach wie vor maßgebend für den Fortbestand und die Budgetierung dieser Abteilungen.<sup>44</sup>

### 3.3 Einsatz von CI im deutschen Mittelstand

Viele mittelständische Unternehmen haben die Bedeutung von Competitive Intelligence für den eigenen unternehmerischen Erfolg noch nicht erkannt. Zwar betreiben nahezu alle mittelständischen Unternehmen CI, jedoch in grundsätzlich unterschiedlicher Ausprägung und Kontinuität. Es findet kaum systematische oder zielgerichtete Konkurrenzbeobachtung statt.<sup>45</sup> Competitive Intelligence bzw. der Erwerb von Kenntnissen über Wettbewerber wird im eigenen Unternehmen vorrangig eingesetzt, um Strategien zu entwickeln und Vorteile für das Unternehmen abzuleiten, Benchmarking durchzuführen und Frühwarnungen vor potenziellen Marktrisiken zu geben. Den Stellenwert des Einsatzes von Competitive Intelligence im Unternehmen schätzten in der Studie der FH Gießen-Friedberg immerhin 28 % der befragten Unternehmen mit „sehr hoch“, 47% mit „hoch“ und noch 20% als „mit-

---

<sup>41</sup> GIA, White Paper, CI in large companies, 2005, S. 11.

<sup>42</sup> GIA, White Paper, CI in large companies, 2005, S. 20.

<sup>43</sup> Schomisch, Michael: Mehr Wissen, innovativer beraten, 2006 S. 46.

<sup>44</sup> Request 2006 – das interne Marketing von CI und InfoDL in Großunternehmen 2006, S. 18-19.

<sup>45</sup> Michaeli/Praetorius: Die neuen Herausforderungen für CI, 2005, S.8.

telwichtig“ ein. Unternehmens-strategische Erkenntnisse durch den Einsatz von Competitive Intelligence werden dabei zumeist aus Verfahren wie die Erstellung von Wettbewerberprofilen, Finanz- und Kennzahlenanalysen, SWOT-Analysen, Szenariotechniken, Gewinn- und Verlustanalysen und Simulationen gewonnen:<sup>46</sup> Dies wurde auch durch die eigene Befragung der Fachhochschule Köln für den Request Wettbewerb 2005 des DCIF (Deutschen Competitive Intelligence Forum) bestätigt.<sup>47</sup>

Idealerweise setzt sich eine CI-Einheit aus einem mit der Branche vertrauten Analysten und einem Informationsspezialisten zusammen, welcher über mehrjährige Erfahrung mit elektronischen Quellen verfügt. Mittelständische Unternehmen verfügen aber in der Regel nicht über diese nötigen Ressourcen für eine eigene CI-Abteilung, so dass dieser Aufgabenbereich unter anderem vom Marketing- oder Marktforschungs-Verantwortlichen übernommen und „nebenbei“ erledigt wird. Bisher werden also ad hoc Recherchen zu spezifischen Sachverhalten durchgeführt und diese dann je nach Größe des Unternehmens direkt von der Geschäftsführung, der Assistenz des Managements oder aber über die Mitarbeiter aus den Bereichen Marketing, Marktforschung oder Vertrieb abgewickelt. Dies hat zur Folge, dass kleine und mittlere Unternehmen häufig strategisch relevante Informationen über Markt, Konkurrenz und Technologien nicht vorausschauend zum Vorteil der Unternehmung einsetzen.<sup>48</sup>

Professionelle und kontinuierlich eingesetzte Wettbewerbsbeobachtung ist daher in der Praxis beim deutschen Mittelstand bisher selten anzutreffen. Gründe hierfür sind oft Zeitmangel und Befürchtungen zu hoher Kosten. Um Zeit und Kosten bei der Auswertung und Analyse komplexerer Sachverhalte zu sparen, könnten externe Dienstleister beauftragt werden, die das nötige Know-how und den Zugang zu verschiedensten Informationsquellen haben.<sup>49</sup> Jedoch könnten für kleine Unternehmen mit festem und stabilem Kundenstamm, die darüber hinaus in einem Markt agieren, der stabile Rahmenbedingungen aufweist, die Kosten den Nutzen übersteigen. Der Einsatz von Competitive Intelligence ist für jedes mittelständische Unternehmen genau abzuwägen. Hierzu sollten marktsegment- und branchenspezifische Untersuchungen durchgeführt werden.

---

<sup>46</sup> Hartwig: Competitive Intelligence, 2003, S.54.

<sup>47</sup> Vgl. Request 2005: Einsatz und Erfolg von Competitive Intelligence im deutschen Mittelstand, 2005.

<sup>48</sup> Vgl. Götte: Competitive Intelligence – denn Wissen ist Macht, 1997.

<sup>49</sup> Vgl. Ullmann: Den Finger am Puls des Marktes, 2005.

### 3.3.1 Informationsbeschaffung und Informationsvermittlung im deutschen Mittelstand

Auch kleine und mittelständische Unternehmen sammeln jeden Tag große Mengen von Daten über ihre Wettbewerbssituation und sind damit wertvolle Wissensträger für erfolgreiches Competitive Intelligence. Wie Shaker und Gembicki in „The War Room Guide to Competitive Intelligence“ deutlich machen, setzt sich dieses Wissen aber zu 80% aus unverarbeiteten Daten (z. B. Verkaufs- oder Messeberichte oder Fachartikel) zusammen, und zu 15% aus „value-added-information“ und lediglich zu 5% aus „intelligence“.<sup>50</sup> Im Rahmen einer Studie aus dem Jahr 2000 über das Informations- und Kommunikationsmanagement in kleinen und mittleren Unternehmen wurden 297 Unternehmen nach ihrem Informationsverhalten und den eingesetzten Informationsquellen und -techniken befragt. Kleine und mittlere Unternehmen nutzen für ihre Informationsbeschaffung vorwiegend direkte persönliche Gespräche und ihre telefonischen Kontakte zu Lieferanten und Kunden, dicht gefolgt von den Printmedien wie Fachzeitschriften und Messen. Öffentliche Informationsquellen wie Behörden, Verbände und Organisationen werden als zuständige Informationsstellen eher in Anspruch genommen als beispielsweise Fachdatenbanken für Marktrecherchen. Bei der Auswahl der eingesetzten Informationstechniken und Nutzung der Informationsquellen werden insbesondere die Beschaffungskosten (90%) und Unterhaltungskosten (77%) berücksichtigt. Erst danach legen die befragten Unternehmen Wert darauf, eine Verbesserung ihrer Informationsbasis (78%) und einen Wettbewerbsvorteil (73%) zu erreichen.<sup>51</sup> Eine Vielzahl von Informationstechniken bleibt also ungenutzt und so ist zu vermuten, dass es vielen kleinen und mittleren Unternehmen an personellen und finanziellen Ressourcen fehlt, um alle verfügbaren Informationsquellen zu identifizieren und zu nutzen, und um so Wettbewerbsvorteile bei der Gewinnung, Analyse und Aufbereitung von externen Informationen zu erreichen. Wichtig ist die Einsicht, dass erst Kontinuität und Systematik in der Wettbewerbsanalyse dazu führt, dass wichtige Ereignisse wie Produkteinführung oder der Markteintritt eines neuen Wettbewerbers nicht verpasst werden.

So wurde im Rahmen von Telefoninterviews des Institutes für Informationswissenschaft der Fachhochschule Köln anlässlich des Request Wettbewerbs 2005 des DCIF festgestellt, dass für den Mittelstand insbesondere Verbandsstrukturen als

---

<sup>50</sup> Vgl. Shaker/Gembicki: The War Room Guide to Competitive Intelligence, 1999.

<sup>51</sup> Teves: Informations- und Kommunikationsmanagement in kleinen und mittleren Unternehmen, 2000, S. 26f.

eine wichtige Informationsquelle angesehen werden, um sowohl auf verbandsinterne Statistiken zuzugreifen als auch Netzwerkstrukturen pflegen zu können. Diese von der Geschäftsführung gewonnenen Informationen werden zusammen mit Kenntnissen aus dem Vertrieb in semi-systematischer Form (z. B. als Aktennotizen) einer zentralen Person oder Stelle im Unternehmen weitergeleitet. Die Qualität der durch Netzwerke gewonnenen Kenntnisse wird generell im deutschen Mittelstand als hoch bzw. vertrauenswürdig, der Informationsgehalt von Pressequellen eher als gering eingeschätzt. Schließlich erfolgt auch die Herausgabe eigener Informationen an die Presse mit verhältnismäßig großer Vorsicht. Bezüglich der von Lieferanten gewonnenen Erkenntnisse ist man sich bewusst, dass diese auch Wettbewerbern zur Verfügung stehen dürften und somit keine Exklusivität besteht. Es zeigt sich zudem, dass die Unternehmen externen Informationsdienstleistern eher kritisch gegenüber stehen, da deren umfangreiche Branchenkenntnis und Erfahrungswissen infrage gestellt werden. Informationsquellen, denen im Mittelstand die größere Bedeutung beigemessen wird, sind also eher informelle Netzwerke. Insbesondere auf Messen, Konferenzen, Workshops oder auch in Telefongesprächen mit Experten, Wissenschaftlern und Praktikern werden Erfahrungen ausgetauscht und neueste Trends weitergegeben, die noch in keiner Datenbank zu finden sind. Dieser „eigene Nachrichtendienst“ ergänzt oder ersetzt die formellen Informationsquellen wie professionelle Datenbanken und Beratungsunternehmen.<sup>52</sup>

#### **4 Fazit**

In Deutschland wird die Informationsvermittlung und der Einsatz von Competitive Intelligence in Großunternehmen und im Mittelstand noch recht unterschiedlich gehandhabt. Insbesondere kleinen und mittelständischen Unternehmen fällt es schwer, Budgets für die systematische Umsetzung von Competitive Intelligence bereitzustellen. Sie schrecken vielfach vor der kontinuierlichen Belastung der fixen Kosten zurück. Doch der Einsatz von Competitive Intelligence ist lohnenswert. In einigen Fällen – wie z. B. bei einer österreichischen Sparkasse, die 1997 in Osteuropa tätig wurde – konnte ein Unternehmen mit einer gut geplanten, CI-gestützten Unternehmensstrategie den Schritt vom mittelständischen zum großen Unternehmen vollziehen.<sup>53</sup>

---

<sup>52</sup> Request 2005: Einsatz und Erfolg von Competitive Intelligence im deutschen Mittelstand, 2005, S. 13-15.

<sup>53</sup> Vgl. Leitl: Competitive Intelligence, 2005.

Die Bedeutung des Einsatzes von Competitive Intelligence und daraus resultierende mögliche Erfolge werden zwar immer mehr vom deutschen Mittelstand erkannt, aber nicht unbedingt umgesetzt. Die Auseinandersetzung mit CI rührt daher, dass mittelständische Unternehmen sich heutzutage dem starken Druck dynamischer Märkte und zunehmender Internationalisierung ausgesetzt sehen, was zu steigender Konkurrenz aber auch zu potentiellen neuen Märkten führen kann. Damit gewinnt die kontinuierliche Beobachtung der eigenen Wettbewerber für den Mittelstand immer mehr an Relevanz, für die Großunternehmen ein schon langer notwendiger Prozess. Einige Methoden und Instrumente von Competitive Intelligence finden im deutschen Mittelstand bereits Anwendung, sie werden jedoch eher unsystematisch und unter Nutzung konventioneller Informationsquellen betrieben. Besonders hervorzuheben ist hier die Nutzung informeller Netzwerke im Mittelstand, das beinhaltet die Einbindung in eine Verbandsstruktur und persönliche Gespräche mit Kunden, Lieferanten und Wettbewerbern. Dabei besteht durchaus Bedarf an einer systematischen Informationsbeschaffung und Informationsvermittlung, die durch externe Dienstleister oder intern durch entsprechende CI-Tools erfolgen könnte. Bei Großunternehmen hingegen wird die strukturierte und prozessorientierte Informationsvermittlung auf Grund der Größe des Unternehmens und der Anzahl an Mitarbeitern eher als Notwendigkeit erkannt und ist finanziell auch eher tragbar. Einige mittelständische Unternehmen fürchten die langwierige Aufbauphase, die notwendig ist, um CI voll funktionsfähig nutzen zu können. Es bestehen zudem Schwierigkeiten, die Erfolge messbar zu machen, da diese zum Teil erst nach Jahren in vollem Umfang offensichtlich werden. Daher könnte der Mittelstand durchaus vom Einsatz externer CI-Experten profitieren, die zum einen eine objektive Beurteilung erstellen zum anderen langfristig keine Personalkosten verursachen. Auch der Mangel an qualifiziertem Personal ist ein weiteres Argument, die Angebote externer Dienstleister zu nutzen und damit Aktivitäten im Bereich Informationsvermittlung einschließlich Competitive Intelligence auszugliedern. Darüber hinaus kann eine Auslagerung der CI-Funktion zu ersten Qualitätsbeurteilungen herangezogen werden, um einen geplanten innerbetrieblichen Einsatz vorzubereiten. Anhand der im Rahmen von Telefoninterviews erstellten Fallbeispiele (Request Wettbewerb 2005 des DCIF) konnte jedoch ermittelt werden, dass der deutsche Mittelstand nicht ausreichend über das Angebot externer Dienstleister für Competitive Intelligence informiert ist. Die Akzeptanz externer Informationsdienstleister durch den deutschen Mittelstand wird aber nur dann erfolgen, wenn die spezifischen Bedürfnisse der mittelständischen Unternehmen erfüllt werden, das heißt wenn eine individuelle und branchenspezifisch ausgerichtete Betreuung erfolgen kann. Auch tiefer gehende Informationen und Analysen über die oft sehr heterogenen Kundenkreise des

Mittelstands bieten weitere Ansatzpunkte für externe Dienstleister.<sup>54</sup> Im Gegensatz dazu stellt sich die Situation in deutschen Großunternehmen anders dar: CI ist bekannt und findet Anwendung, wenn auch nicht immer systematisch und unternehmensübergreifend. Oft sind es gleich mehrere Abteilungen im Unternehmen, die sich mit der Wettbewerbsanalyse und CI befassen, da diese nicht nur in der Informationsabteilung erfolgt. Dadurch erschwert sich die Koordination und Durchführung von CI und dezentralisierte Abteilungen laufen Gefahr auf sich allein gestellt zu werden. Nur wenige der für den Request Wettbewerb 2006 befragten Großunternehmen hatten CI-Abteilungen, die ihre Leistungen allen Abteilungen anbieten. Der überwiegende Teil bietet die eigene Dienstleistung und Ergebnisse nur bestimmten Nutzerkreisen an, in den meisten Fällen dem Management bzw. der Geschäftsführung. Vereinzelt gibt es einzelne CI-Unterabteilungen als Teil von größeren Abteilungen, wie zum Beispiel dem Marketing, die ihre Leistungen auch nur ihren eigenen Abteilungen zur Verfügung stellten.<sup>55</sup> Eine wichtige Rolle spielt auch die Einbindung in den unternehmerischen Prozess, um die Akzeptanz der Leistungen dieser Informationsabteilungen im gesamten Unternehmen zu gewährleisten. Ein befragtes Großunternehmen versucht beispielsweise seit einigen Jahren eine Unternehmenskultur zu schaffen, die alle Abteilungen des Unternehmens von der Wichtigkeit und Bedeutung dieser Abteilung (einschließlich CI) für den Gesamtunternehmenserfolg überzeugt und die eine sinnvolle abteilungsübergreifende Zusammenarbeit möglich machen soll. Die Höhe des Budgets und die Anzahl der Mitarbeiter spielen insofern eine wichtige Rolle, dass bei Überlastung der Kapazitäten viele Anfragen nicht zeitnah bearbeitet werden können. Auch spezielle Fragestellungen sind so zeit- und kostenintensiv, dass die angestrebte hohe Qualität der Ergebnisse nicht immer gewährleistet werden kann. In diesen Fällen werden externe Informationsvermittler / Dienstleister von den Großunternehmen in Anspruch genommen. Auch im Rahmen von Großprojekten wurden in der Vergangenheit in Großunternehmen Freelancer (feste Unternehmenspartner) zur Unterstützung dieser Tätigkeiten herangezogen.

Auslagerung oder Outsourcing von CI und anderen Informationsdienstleistungen ins Ausland wurde von allen Befragten als kritisch und schwer durchführbar eingestuft. Viele Befragte aus den Großunternehmen sehen darin eine Gefahr der Sicherheit der strategischen Informationen und mögliche Qualitätseinbußen. Trans-

---

<sup>54</sup> Request 2005: Einsatz und Erfolg von Competitive Intelligence im deutschen Mittelstand, 2005, S.18-20.

<sup>55</sup> Vgl. Request 2006: Das interne Marketing von CI und InfoDL in Großunternehmen 2006.

ationale Unternehmen stufen hingegen die CI-Arbeit unter Zusammenarbeit mit lokalen Partnern in den Ländern in denen sie tätig sind als weniger gefährdend sondern vorteilhaft ein, wenn Aufgaben und Untersuchungsbereich klar umrissen sind, dies gilt vor allem für den asiatisch-pazifischen Raum. Outsourcing ist dann möglich, wenn die Dienstleistung als solche keine Fachexpertise erfordert, so wird diese Art des outsourcings schon bei vier der befragten Unternehmen eingesetzt.<sup>56</sup> Gerade im asiatischen Raum (Indien) werden CI-Dienstleistungen im Rahmen von Research und Beratung (CI eingeschlossen) vermehrt und kostengünstig angeboten. Weiterhin ist davon auszugehen, dass CI in Zukunft an Bedeutung gewinnen wird, Schätzungen prognostizieren die weltweiten Ausgaben im Bereich CI im Jahre 2010 auf über 400.000.000\$. Outsourcing könnte hier Kosteneinsparungen von 30 – 40 % realisieren. Es wird erwartet, dass zunächst Unternehmen in den USA und Großbritannien hier eine Vorreiterrolle einnehmen werden.<sup>57</sup>

## 5 Literatur

- Altensen, Astrid / Pfaff, Dietmar [2006]: : Fachhochschule Gießen-Friedberg: Ergebnisse einer bundesweiten Studie zur Competitive Intelligence. In: Michaeli, R.: Competitive Intelligence – Strategische Wettbewerbsvorteile erzielen durch systematische Konkurrenz-, Markt- und Technologieanalysen, Berlin: 2006, S.60-82
- Bamberger, Ingolf/ Evers, Michael [2002]: Ursachen und Verläufe von Internationalisierungsentscheidungen mittelständischer Unternehmen. In: Macharzina, K.: Handbuch Internationales Management, 2. Aufl. Wiesbaden 2002, S. 273-313
- Brellochs, Andreas [2000]: ‚Competitive Intelligence‘ und ‚Knowledge Management‘, Dipl.Arb. Konstanz 2000
- Bruhn, Manfred [1999]: Internes Marketing 2. Aufl., Wiesbaden, 1999
- CI-Handbuch [2005]: Stuttgart, <http://www.ci-handbuch.de/bt000020.htm> (14.10.2006)
- Deltl, Johannes [2004]: Strategische Wettbewerbsbeobachtung, Wiesbaden 2004
- DCIF [2005]: Was ist Competitive Intelligence? München <http://www.dcif.de/ci.html> (30.10.2006)
- GIA (Global Intelligence Alliance) (Hrsg.) [2005]: White Paper 4/2005: Competitive Intelligence in large companies – global study
- Götte, B. [1997]: Competitive Intelligence – denn Wissen ist Macht. In: ioManagement 12(1997), S. 40-46

---

<sup>56</sup> Vgl. Request 2006: Das interne Marketing von CI und InfoDL in Großunternehmen 2006.

<sup>57</sup> Vgl. Studie „Emerging Trends in Market Intelligence Outsourcing“, SCOPE eknowledge center, 2006.



- Hartwig, Stefan [2003]: Competitive Intelligence – vertane Chance und vernachlässigtes Risiko. In: Wissensmanagement, 7 (2003), S.53
- Hoppenstedt (Hrsg.) [2006]: Pressemitteilung zu Großunternehmen vom 20. 4. 2006, [http://www.fastx.de/index.htm?0503/archiv2006/FI-GU\\_06-1-PI-E.htm](http://www.fastx.de/index.htm?0503/archiv2006/FI-GU_06-1-PI-E.htm) (3.4.2007)
- Institut für Mittelstandsforschung (Hrsg.) [2003], Bonn: Definitionen und Schlüsselzahlen, <http://www.ifm-bonn.org/index.htm?/presse/mind-03.htm>
- Kayser, Gunter [2005]: Institut für Mittelstandsforschung Bonn: Was sind eigentlich kleine und mittlere Unternehmen (KMU)? Bonn, <http://www.ifm-bonn.org/presse/kay-aif.pdf> (5.4.2007)
- Korn-Ferry International/University of Southern California [2000]: Strategies for the Knowledge Economy, Los Angeles, USA 2000
- Leidl, Michael [2005]: Competitive Intelligence? In: Harvard Business Manager. 5 (2005) S. 14
- Lux, Christian / Nösner, Axel [2004]: Competitive Intelligence erfolgreich anwenden, In: Wissensmanagement, 5 (2004), S. 19-20
- Pfaff, Dietmar/Altensen, Astrid: Competitive Intelligence – Nur eine Domäne der Amerikaner.? In: Absatzwirtschaft online [http://www.absatzwirtschaft.de/Content/\\_pv/\\_p/1003378/\\_t/fthighlight/highlightkey/pfaff+altensen/\\_b/35074/default.aspx/nur-eine-domaene-der-amerikaner.html](http://www.absatzwirtschaft.de/Content/_pv/_p/1003378/_t/fthighlight/highlightkey/pfaff+altensen/_b/35074/default.aspx/nur-eine-domaene-der-amerikaner.html) (3.4.2007)
- Pfaff, Dietmar [2005]: Competitive Intelligence in der Praxis, Frankfurt a.M., 2005
- Pfaff, Dietmar / Glasbrenner, Claudia [2004]: Nachholbedarf in Sachen Competitive Intelligence. In: Wissensmanagement (2004) Nr. 3. S. 43-45
- Reichwald, Ralf / Bruhn, Manfred [2005]: Führung, Organisation und Kommunikation, In: Zfo Zeitschrift Führung + Organisation 3 (2005) S. 132-138
- ReQuest 2005: Fachhochschule Köln- Institut für Informationswissenschaft: Einsatz und Erfolg von CI im deutschen Mittelstand. FH Köln anlässlich der DCIF Tagung in Andechs, 2005
- ReQuest 2006: Fachhochschule Köln- Institut für Informationswissenschaft: Das interne Marketing von CI und Informationsdienstleistungen in Großunternehmen. FH Köln anlässlich der DCIF Tagung in Dresden, 2006
- Michaeli, Rainer [2006]: Competitive Intelligence, – Strategische Wettbewerbsvorteile erzielen durch systematische Konkurrenz-, Markt- und Technologieanalysen, Berlin: 2006
- Michaeli, Rainer/Praetorius, Jörg [2005]: Die neuen Herausforderungen für Competitive Intelligence, In: Password 2 (2005), S. 8-9
- Schomisch, Michael [2006]: Mehr Wissen, innovativer beraten, In: Wissensmanagement, (2006) H. 2, S. 46
- SCOPE eknowledge center, Studie [2006]: „Emerging Trends in Market Intelligence Outsourcing“, September 2006
- Seufert, Andreas und Lehmann, Peter [2005]: Business Intelligence – Status quo und zukünftige Entwicklungen, In: HMD 2005, S. 21-32

- Shaker, Steven M. / Gembicki, Mark P. [1999]: The War Room Guide to Competitive Intelligence. New York, 1999
- Statistisches Bundesamt, Wiesbaden,  
<http://www.destatis.de/basis/d/insol/unternehtab3.php> und  
<http://www.destatis.de/basis/d/insol/unternehtab2.php> (8.4.2007)
- Stippel, Peter [2002]: Fast wie beim CIA – Konkurrenzabwehr im globalen Wettbewerb, In: Absatzwirtschaft 45 (2002) H.4, S. 14-20
- Teves, Nikolaus [2000]: Informations- und Kommunikationsmanagement in kleinen und mittleren Unternehmen (Veröffentlichungen des Instituts für Mittelstandsforschung, 41), Mannheim, 2000
- Ullmann [2005]: Den Finger am Puls des Marktes, In: ProFirma 5 (2005), S. 14-17
- Wikipedia, Großunternehmen: <http://de.wikipedia.org/wiki/Gro%C3%9Funternehmen> (8.4.2007)

# Weblogs als Steuerungsinstrument in Hochschulbibliotheken

## Optimierung von Dienstleistungen an Hochschulbibliotheken auf der Basis von Web 2.0 Technologien\*

*Christine Krätzsich*

Universität Mannheim

Universitätsbibliothek

Schloss Ostflügel

68131 Mannheim

*christine.kraetzsch@bib.uni-mannheim.de*

### **Zusammenfassung**

Die Universitätsbibliothek Mannheim versteht sich als moderne Dienstleistungseinrichtung, die ihre Geschäftsprozesse unter Berücksichtigung einer stärkeren Kundenorientierung neu ausrichten will. Sie möchte die neuen Perspektiven einer sich verändernden Partizipations- und Kommunikationskultur im Kontext von Web 2.0 nutzen, um ihr Informations- und Steuerungsmanagement stärker interaktiv zu gestalten. Im Rahmen eines DFG-Projekts untersucht sie derzeit, inwieweit sich Web 2.0-Anwendungen für diese Zwecke einsetzen lassen.

### **Abstract**

The library of the University of Mannheim understands itself as modern service orientated institution, which wants to revise their business processes with consideration of a stronger customer orientation. It should be taken note of the new perspectives of a changing participation and communication culture in the context of Web 2.0, in order to arrange the information and control management more strongly interactive. In the context of a research project which is supported by the Deutsche Forschungsgemeinschaft (DFG) it examines at present, to what extent Web 2.0 applications can be used for these purposes.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 207-220.

## I Veränderte Lebenswirklichkeit 2.0

Im Kontext der Weiterentwicklung des World Wide Web konstituieren sich unter dem Schlagwort Web 2.0 nicht nur eine Vielzahl von neuen interaktiven Services, sondern vor allem neue Kommunikations- und Partizipationsformen, die über das Internet hinaus die Lebenswirklichkeit nachhaltig verändern.

Zu den bekanntesten Web 2.0 Anwendungen gehören Weblogs, die nicht nur als private Online-Tagebücher, sondern vor allem als fachspezifische Diskussions- und Newsplattformen eine wachsende Bedeutung und Rezeption erlangt haben. Was private oder kommerzielle Homepages mit ihren täglich oder stündlich wechselnden Inhalten nur unzureichend vermochten, da sie wenig nachvollziehbar und in ihrer Masse unüberschaubar blieben, können Blogs heute viel besser: Informationen aktuell und chronologisch greifbar und über Kategorisierungen benutzerfreundlich zugänglich machen. Es sind verhältnismäßig kleintechnologische Innovationen wie RSS<sup>1</sup> und Permalink<sup>2</sup>, die eine vollkommen neue Art der Vernetzung und des webbasierten Dialogs ermöglicht haben.

Auf Basis dieser Echtzeit-Vernetzung gewinnt das Internet an Realität. Der Weg der Kommunikation mag virtuell sein, die Kommunikation selbst findet real, zwischen einer großen und stetig wachsenden Gruppe von Menschen statt. Es ist das eigentlich Spannende am Phänomen Web 2.0, dass es sich nicht schlicht auf eine technische Innovation reduzieren lässt. Wenn mittlerweile mehr als eine Milliarde Menschen online sind (vgl. Kleinwächter 2007), die das Netz zunehmend selbst gestalten, indem sie Videos, Kommentare, Ideen etc. einstellen und auf der Basis von Open Source die Anwendungen ihren individuellen Bedürfnissen anpassen und weiterentwickeln, dann verändern sie nicht nur das Internet, sondern über die Kommunikation das Leben selbst.

Populäre Weblogs wie *Bildblog.de* verzeichnen pro Tag Seitenabrufe im fünfstelligen Bereich (vgl. *blogscout.de*), einzelne Videos auf *YouTube* werden sogar millionenfach angeklickt (vgl. *youtube.com*). Als dialogorientierte Informations- und Kommunikationsplattformen etablieren sich einige Web 2.0 Services zunehmend

---

<sup>1</sup> Mit Hilfe von RSS (Real Simple Syndication, auch Rich Site Summery) können Webseiten und ihre Leser in Echtzeit mit Informationen versorgt und über Änderungen informiert werden.

<sup>2</sup> Permalink ist die Bezeichnung für eine permanente Verknüpfung mittels derer jeder Blog (Eintrag, Kommentar, etc.) über eine eindeutige URL zeitunabhängig abrufbar ist.

als ernsthafte Konkurrenz zu den klassischen Medien. Zerfass (2006) spricht in diesem Zusammenhang von dem Beginn einer neuen Ära gesellschaftlicher Kommunikation, die ihre Wirkung auch über das Internet hinaus auch in Wirtschaft und Gesellschaft entfalten wird (vgl. auch Schroll et al. 2007).

Besonderes Kennzeichen dieser neuen Kommunikationswelt ist die Erosion der starren Grenze zwischen Autor und Rezipient. Im Web 2.0 werden sowohl die Inhalte als auch die Formen der Anwendungen in mehr oder weniger freien Kooperationen von den Nutzern selbst gestaltet. Auch Tim O'Reilly, der den Begriff Web 2.0 als Erster verwendet hat<sup>3</sup>, erachtet den Gedanken der Plattform, auf der ein freier Austausch zwischen Menschen mit gleichen Interessen stattfinden kann, als zentral für das Web der Zukunft (vgl. O'Reilly 2005; Lange 2006). Grundlage dieser Entwicklung sind Open Source Anwendungen, die von jedem kostenlos genutzt und weiterentwickelt werden können und deren Verbesserungen dann wieder jedem kostenlos zugänglich sind. Im Web 2.0 wird Mitmachen Mainstream, das Internet zum Mitmach-Web. In der Folge ergibt sich der interessante Effekt, dass die Erstellung von Software durch eine praktisch unendlich große Zahl von Entwicklern, Anwendungen hervorbringt, die den Bedürfnissen der Nutzer viel besser gerecht werden können als herkömmliche lizenzierte Software.

## **2 Online-Marketing und Corporate Blogging**

Wenn Web 2.0 Services in Konkurrenz zu den klassischen Medien an Einfluss gewinnen und darüber hinaus der User immer mehr zu der zentralen Größe im Netz wird (vgl. Schroll et al. 2007), muss auch in den Marketingabteilungen von Unternehmen ein Nachdenken über Web 2.0 einsetzen.

Firmen wie Amazon oder guentstiger.de machen schon länger vor, wie die Kommunikations- und Informationsbedürfnisse der Kunden mit Hilfe von Web 2.0 Features bedient und zugleich für eigene Zwecke genutzt werden können. Sie stellen innerhalb der eigenen Angebote Plattformen zur Verfügung, auf denen sich die Kunden gegenseitig über die angebotenen Produkte austauschen können. Die dort (nahezu kostenfrei) gesammelten Informationen sind in zweierlei Hinsicht sehr wertvoll. Zum einen bieten sie weiteren Kunden sinnvolle Hilfestellung bei der Auswahl der passenden Produkte und sorgen schon auf dieser Ebene für mehr Kundenzufriedenheit. Zum anderen stellen die Erfahrungen und Bewertungen der

---

<sup>3</sup> 2004 zusammen mit Dale Dougherty im Rahmen eines Brainstormings zu Trends im Web. Vgl. Lange, Corina (2006).

Kunden eine mächtige Ressource dar, die für die Optimierung der eigenen Steuerungsprozesse genutzt werden kann.

Durch die Auswertung der von Kunden generierten Daten kann nicht nur ersichtlich werden, welche Produkte bei den Kunden gut ankommen und welche Kunden welche Produkte bevorzugen. Mit Hilfe statistischer Verfahren können Kundenprofile erstellt und Marketingaktionen im Sinne von One-to-One-Marketing individuell auf einzelne Kunden(-gruppen) zugeschnitten werden. Darüber hinaus können die Daten als Grundlage für eine konsequente Individualisierung des Angebots, die Kostensenkung von Geschäftsprozessen und eine Vergrößerung der Reichweite durch Cross-Media-Marketing genutzt werden (vgl. Conrady et al. 2002).

Für Online-Marketing-Strategien eröffnen die informationstechnologischen Veränderungen um Web 2.0 neue Kommunikations- und Distributionsmöglichkeiten, mit denen eine verstärkte Hinwendung zum Kunden erreicht werden kann. Das kurzfristig angelegte und auf Produkt und Leistung fokussierte Transaktionsmarketing mit dem Primärziel der Kundenakquisition wird mehr und mehr durch netzwerkbasierendes, individuelles und multioptionales Beziehungsmarketing abgelöst werden (vgl. Huber 2006). In Ergänzung zum klassischen Marketing-Mix ermöglichen Web 2.0 Anwendungen durch ihre starke Dialogorientierung eine effektivere Kommunikation mit den Kunden und in der Folge mehr Kundenzufriedenheit und langfristige Kundenbindung.

Um diese Chancen optimal nutzen zu können, sind Unternehmen zunächst gefordert, sich mit den neuen Kommunikationsformen und den konkreten Lebenswirklichkeiten ihrer potentiellen Kunden auseinanderzusetzen. Welche Auswirkungen hat eine sich verändernde öffentliche Meinungsbildung für das unternehmenseigene Kommunikationsmanagement? Inwiefern müssen die bisherigen Strategien angepasst und neue Maßnahmen geplant werden – vor allem auch im Zusammenhang mit den angebotenen Produkten bzw. Dienstleistungen und der anvisierten Zielgruppe? Welche Web 2.0 Anwendungen eignen sich für welches Unternehmensziel ganz konkret? Schließlich muss überlegt werden, wie die Kunden zu einer freiwilligen und sowohl quantitativ als auch qualitativ ergiebigen Beteiligung im Sinne der Bereitstellung von relevanten Informationen bewegt werden können.

Welche Potentiale, Herausforderungen und Risiken sich z. B. mit Weblogs in der Unternehmenskommunikation verbinden, wird bereits seit einiger Zeit untersucht (vgl. z. B. Berlecon Kurzstudie 2004; Zerfass 2005; Schmidt 2006; Schroll et al. 2007). Einig sind sich die Analysten vor allem in einem: Ignorieren können Unter-

nehmen dieses neue Medium nicht. Es ist jedoch kontraproduktiv mit herkömmlichem Marketingvokabular und -instrumentarium zu arbeiten, da herkömmliche Werbestrategien unter den Vorzeichen einer veränderten Kommunikationsweise nicht zwangsläufig funktionieren.

Wenn die Spezifika von Weblogs beachtet wird, können Unternehmens-Weblogs (Corporate Blogs)<sup>4</sup> jedoch sehr sinnvoll – sowohl intern als auch extern – für eine effizientere Kommunikation der Mitarbeiter untereinander oder mit den Kunden eingesetzt werden. Vor dem Start eines Blogs muss jedoch geklärt werden, inwieweit Unternehmensangebot und Kommunikationsform zusammenpassen, in welcher Weise ein Blog die Kundenbindung und -loyalität unterstützen kann und welche Mitarbeiter geeignet sind, ein Weblog zu betreiben. Dabei sollte insbesondere auf die Konsistenz mit anderen Kommunikationskanälen, die Benutzerfreundlichkeit, die zeitnahe und konstruktive Reaktion auf Kritik und auf die Authentizität der Schreibenden geachtet werden.

### **3 Herausforderungen für Bibliotheken**

Im Zuge aktueller Umstrukturierungsprozesse haben sich die Bibliotheken bisher ausschließlich auf die Erfassung der Betriebskosten und mögliche Rationalisierungseffekte im Rahmen der Kosten-Leistungs-Rechnung konzentriert. Bei der Untersuchung der spezifischen Wünsche der Kunden legen die Bibliotheken ihren Schwerpunkt auf die Erfassung von Kennzahlen und die Übertragung des Modells der Balanced Scorecard. Ein großes Problem bei der Anwendung dieses Modells liegt darin, dass zwar die Perspektive der Kunden berücksichtigt wird, damit jedoch noch kein verlässliches Instrument zur Verfügung steht, um zu messen, inwieweit die Kunden mit den Dienstleistungen der Bibliothek zufrieden sind.

Vor dem Hintergrund der Entwicklungen um Web 2.0 empfiehlt es sich auch für Bibliotheken, die Bedürfnisse ihrer Kunden stärker als bisher in die strategische

---

<sup>4</sup> Eine Definition von Corporate Blogs gibt Zerfass: "Corporate Blogs (bzw. Unternehmens-Weblogs) sind persönliche oder thematische Nachrichtendienste, die von Unternehmen oder mit deren Unterstützung von Mitarbeitern unter Verwendung einfacher Content Management Systeme als Website im Internet publiziert, in regelmäßigen Abständen ähnlich wie ein Tagebuch um neue Einträge ergänzt und in vielfältiger Weise mit anderen Blogs und Websites verlinkt sind. Die Einträge sind umgekehrt chronologisch angeordnet und können jederzeit von den Nutzern kommentiert sowie mit Hilfe normierter Technologien in Echtzeit in andere Informationsangebote eingebunden [...] werden" (2005: 3, Herv. i. O.).

Planung einzubeziehen. Das gilt auch und besonders für Hochschulbibliotheken. Deren Hauptklientel wird zunehmend mit dem Internet und Web 2.0-Anwendungen aufgewachsen und vertraut sein. Die Selbstverständlichkeit mit der sich nachwachsende Studierendengenerationen in virtuellen Räumen bewegen, wird sich voraussichtlich auch in geänderten Erwartungshaltungen an Einrichtungen wie Hochschulbibliotheken widerspiegeln. Nur wenn Bibliotheken in Antizipation dieser Ansprüche anfangen zu agieren statt zeitverzögert – und daher z. T. an den aktuellen Bedürfnissen vorbei – zu *reagieren*, werden sie in der Lage sein, sich in einer Wissensgesellschaft, deren Informations- und Kommunikationsstrukturen sich derzeit grundlegend wandeln, zu behaupten (vgl. Halle 2005).

Kommerzielle Anbieter wie Google und Amazon bauen mit ihren benutzerfreundlichen, schnellen und universalen Informationszugängen ihren Marktanteil bei der Informationssuche zu Lasten von Bibliotheken weiter aus. Im Zuge der Digitalisierung von Informationen laufen Bibliotheken Gefahr auch in Bezug auf weitere Teile ihres Kerngeschäfts, bei der Beschaffung und Erschließung von Informationen, an Bedeutung zu verlieren. Die Folge ist eine Trivialisierung des Informationszugriffs und eine Beschränkung auf die diesen Angeboten zugrunde liegenden Quellen. Die Informationssuchenden sind sich dessen indes kaum bewusst. Mehr oder weniger unbewusst verzichten sie auf bibliothekarisches Wissen und die Nutzung von qualitativ höherwertigen Angeboten.

Um diesen Entwicklungen entgegen zu wirken, sind Bibliotheken gefordert, die bereits seit längerer Zeit anempfohlene Weiterentwicklung von Hol- zu Bringbibliotheken (vgl. WIR 2001) konsequent umzusetzen. Darüber hinaus müssen Bibliotheken ihr Angebot an neuen Ansprüchen ausrichten und dabei zugleich den Mehrwert ihrer bibliothekarischen Kompetenzen aufzeigen und positionieren. Gelingen kann dies nur, wenn Hochschulbibliotheken dazu übergehen, den individuellen Kunden mit seinen ganz konkreten Bedürfnissen ins Zentrum ihrer Angebote zu stellen.

Dafür sind Bibliotheken jedoch auf Informationen über ihre Kunden angewiesen. Wie stellen sich diese eine benutzerfreundliche Bibliothek vor? Was genau wollen sie finden? Welche Techniken nutzen sie am liebsten? Welche Anforderungen an Geschwindigkeit, Bequemlichkeit und Service haben sie? Wie ist es um die Informationskompetenzen der eigenen Kunden bestellt? Welche anderen Services ziehen sie den Angeboten der Bibliothek vor?



## **4 Weblogs als Steuerungsinstrument in Hochschulbibliotheken**

### **4.1 Projektziele**

Wesentliches Ziel des von der DFG geförderten Projekts "Weblogs als Steuerungsinstrument in Hochschulbibliotheken" ist die Entwicklung eines Steuerungsinstruments, mit dessen Hilfe Hochschulbibliotheken Kundenzufriedenheit optimieren und die eigene Organisation auf dieses Ziel ausrichten können. Um dieses Ziel zu erreichen, ist es notwendig, einen Dialog mit den Kunden der Bibliothek anzustoßen und sie dazu zu bewegen, ihre konkreten Bedürfnisse und Ansprüche gegenüber der Bibliothek zu formulieren. Zu diesem Zweck erweitert die Universitätsbibliothek Mannheim ihren Online-Katalog um die Möglichkeit, Medien zu rezensieren und zu bewerten.

Innerhalb des Online-Katalogs bilden die Rezensionen für die Kunden eine Plattform für die Kommunikation untereinander und mit der Bibliothek. Die Kunden erstellen kooperativ Kommentare zu den gedruckten und elektronischen Medien der Bibliothek. Beispielsweise beschreibt ein Student, ob die Lektüre eines bestimmten Lehrbuchs für eine Prüfung hilfreich war. Andere Studierende können diese Empfehlung als hilfreich, nicht hilfreich oder unzumutbar bewerten. Auf diese Weise können Meinungen gewichtet und extreme Äußerungen nivelliert werden. Die Einträge werden mit den Titelaufnahmen im Online-Katalog verknüpft und bieten eine zusätzliche Möglichkeit, die in einer Hochschulbibliothek angebotenen Informationen inhaltlich zu erschließen.

Ein weiterer wichtiger Aspekt ist die Möglichkeit der Kommunikation der Kunden mit der Bibliothek. Sind die Kunden mit dem Angebot der Bibliothek unzufrieden oder haben Verbesserungsvorschläge, haben sie keine Möglichkeit, die Kritik an der Stelle zu äußern, an der die Dienstleistungen angeboten werden. Stellen Studierende z. B. bei einer Katalog-Recherche fest, dass ein gewünschtes Buch über einen längeren Zeitraum vorgemerkt ist und nur einmal in der Bibliothek vorhanden ist, können sie beispielsweise die Bedeutung des Buches für ein Fach erläutern und den Wunsch nach weiteren Exemplaren direkt auf der Katalogseite äußern.

Die Referenten können in ihren Kaufentscheidungen nur bedingt auf die Kundeninteressen eingehen, da ihnen die dazu notwendigen empirischen Daten fehlen. Bisher sind sie darauf angewiesen, dass die Kunden ihnen ihre Wünsche mitteilen. Im Rahmen des Projekts sollen die Einträge automatisiert ausgewertet werden, um zusätzliche Entscheidungshilfe für den Kauf zu geben.

Schließlich soll untersucht werden, ob die konsequente Ausrichtung nach den Kundeninteressen eine neue Organisationsform für eine Hochschulbibliothek erfordert. In diesem Zusammenhang ist denkbar, dass Mitarbeiter vermehrt Aufgaben im Bereich Kundenbetreuung übernehmen und der Anteil der Beschäftigten im Bereich Medienerwerbung und Bestandserschließung verringert wird.

#### **4.2 Vergleich Weblogs und Web 2.0 Service der UB Mannheim**

Weblogs bieten sich als Instrumente eines kundenorientierten Dialogs an. Nicht nur von Unternehmen, sondern auch von Bibliotheken werden sie zunehmend eingesetzt, um die Kommunikation mit den Kunden direkter und bedarfsorientierter gestalten zu können.

Auf der Grundlage von Weblog Software wird es allen Angehörigen der Universität Mannheim ermöglicht, die Medien der UB direkt im Online-Katalog zu rezensieren und zu bewerten<sup>5</sup>. Auf diese Weise können zum einen den Kunden der Bibliothek zusätzliche Hinweise – z. B. hinsichtlich der Relevanz für wissenschaftliche Fragestellungen oder die Eignung für bestimmte Prüfungen – direkt von denjenigen gegeben werden, die die Medien tatsächlich nutzen. Im Hinblick auf die eigenen Geschäftsprozesse erhält die Bibliothek auf der anderen Seite wertvolle Informationen darüber, welche Titel mit welchem Nutzen nachgefragt und bewertet werden.

In Anlehnung an die Funktionalitäten von Weblogs können von allen Studierenden und Beschäftigten der Universität Mannheim Einträge verfasst und gelesen werden<sup>6</sup>, die dann in umgekehrt chronologischer Reihenfolge eingetragen werden. Im Online-Katalog wurde zu diesem Zweck ein Content-Management-System integriert, das eine Eingabe direkt auf der Katalog-Seite eines jeden Titels gestattet. Diese Zusatzinformationen sind – sofern sie von Kunden der Bibliothek eingetragen wurden – direkt mit den Titeln im Online-Katalog verknüpft und somit bei jeder Katalogrecherche verfügbar. Die eingegebenen Rezensionen können von allen Universitätsangehörigen hinsichtlich ihres Nutzens als hilfreich oder nicht hilfreich bewertet werden. Sie bleiben – soweit sie den Richtlinien (vgl. Rezensionsrichtlinien 2007) entsprechen – dauerhaft abrufbar und über eine eigene URL identifizierbar. Es können alle Rezensionen einer Person angezeigt werden.

---

<sup>5</sup> Die Medien können mit Sternen von 5 (sehr empfehlenswert) bis 1 (nicht empfehlenswert) bewertet werden.

<sup>6</sup> Die eingegebenen Rezensionen können ohne Einschränkungen von allen Nutzerinnen und Nutzern des Online-Katalogs gelesen werden. Für den Eintrag einer Rezension ist eine Authentifizierung mittels einer universitätsinternen Benutzerkennung notwendig.

Auf einige Features von Weblogs wurde im Hinblick auf die Projektziele verzichtet. So ist die freie Kommentierung der einzelnen Einträge nicht möglich, ebenso wenig die Verlinkung zu anderen Webseiten. Darüber hinaus werden eine kategorienbasierte Archivierung, RSS-Feeds, Trackbacks und die Nutzung von Bookmarkdiensten zunächst nicht angeboten. Zusätzlich gibt es die Möglichkeit, die Administration per E-Mailformular über Rezensionen, die gegen die Richtlinien verstoßen zu informieren. Dies soll vor allem die Selbstkontrolle der Kunden unterstützen und zur Vermeidung extremer und unerwünschter Äußerungen beitragen.

Die höchste Reichweite und somit den größten Nutzen für die Kunden erreichen die Rezensionen im Online-Katalog. Deshalb wurde sich im Projektverlauf dafür entschieden, anstelle eines Weblogs im engeren Sinne, eine Rezensionsplattform auf der Basis von Web 2.0 Technologien direkt im Online-Katalog zu integrieren.

#### **4.3 Übertragung von Online-Marketing-Konzepten auf den Non-Profit-Bereich**

Im Verlauf des Projekts soll geprüft werden, ob und im welchen Umfang Konzepte aus der Privatwirtschaft auf den öffentlichen Bereich übertragen werden können. Dabei muss zunächst berücksichtigt werden, dass Privatwirtschaft und öffentlicher Sektor auf unterschiedliche Ziele ausgerichtet sind. Unternehmen sind in der Regel auf Wachstum und Gewinnmaximierung ausgerichtet. Im staatlichen Bereich geht es darum, unter Einsatz von begrenzten Finanz- und Personalressourcen eine vom Unterhaltsträger definierte "Dienstleistung" zu erbringen. Der Einsatz zusätzlicher Mittel durch Steigerung des Umsatzes ist hier nur begrenzt vorgesehen.

Für das Beispiel Hochschulbibliothek bedeutet dies, dass diese Einrichtung den zugewiesenen Etat für die Versorgung der Hochschule mit gedruckten und elektronischen Informationen einsetzt. Trotz dieses Unterschieds ist der Einsatz von betriebswirtschaftlichen Steuerungsinstrumenten im öffentlichen Sektor sinnvoll, da Steuerungs- und Überwachungsprobleme zu Ineffizienzen sowie Fehlallokationen der Ressourcen führen (vgl. Schwarz 2001). Dazu kommen die Etatprobleme der öffentlichen Haushalte, die eine effiziente Verwaltung und Einsatz der zugewiesenen Mittel erfordern.

Eine weitere Frage besteht darin, in welchem Umfang die Begriffe Konsument und Kunde auf den öffentlichen Bereich übertragen werden können. Ein wesentlicher Unterschied ist, dass Studierende in der Regel nicht zwischen mehreren Anbietern wählen können, die die Dienstleistung 'Versorgung mit gedruckten und elektronischen Medien' anbieten – die gleichzeitige Benutzung mehrerer Hochschulbibliotheken in Ballungsräumen soll hierbei ausgeklammert werden. Dennoch ist die Ver-

wendung der Begriffe Kunde und Konsument in diesem Zusammenhang sinnvoll, da damit die Bereitschaft zu qualitativ höheren und umfangreicheren Dienstleistungen ausgedrückt wird als dies mit dem Begriff Nutzer möglich ist, der die Rolle des Bittstellers gegenüber der staatlichen Institution impliziert. Durch den Einsatz interaktiver Features können Hochschulbibliotheken die Kommunikation mit ihren Kunden verbessern und eine höhere Marktdurchdringung erreichen, da sie kaum bekannte und wenig genutzte Dienstleistungen platzieren können. Darüber hinaus können die gewonnenen Daten im Sinne einer effizienteren Steuerung des Erwerbsetats und der Kundendienste ausgewertet werden.

#### **4.4 Evaluierung der Kundeneinträge**

Die von den Kunden eingetragenen Rezensionen sollen teilautomatisch mit Hilfe statistischer Verfahren in Anlehnung an Data Mining Methoden ausgewertet werden. Ziel ist es, die gesammelten Datenmengen im Hinblick auf Regeln und Muster bzw. statistische Auffälligkeiten sowie mögliche Änderungen im Verhalten der Kunden oder ganzer Kundengruppen zu untersuchen.

In einem ersten Schritt werden die Datenquellen für die Auswertung ausgewählt: die Einträge der Kunden im Online-Katalog, die Bewertung der Medien und die administrativen Daten im Lokalsystem der Universitätsbibliothek Mannheim (u a. Katalogseitenaufrufe, Anzahl der Vormerkungen pro Exemplar, RVK-Notationen). Anschließend müssen Fragen für eine sinnvolle Auswertung der Daten formuliert werden und ein geeignetes Analyseraster entworfen werden.

Ein wichtiges Interesse liegt in der Ermittlung von Teildisziplinen eines wissenschaftlichen Faches, die für die Lehre eine wichtige Rolle spielen. Hierfür ist eine wichtige Grundlage, dass ein großer Teil des Bestandes der Universitätsbibliothek Mannheim intellektuell und automatisiert mit Hilfe der Regensburger Verbundklassifikation (RVK) systematisiert ist. Die bewerteten Medien können den entsprechenden Gruppen der RVK zugeordnet werden. Auf diese Weise kann überprüft werden, in welchen Teildisziplinen besonders viele Rezensionen eingetragen werden und welche durchschnittlichen Bewertungen dabei vergeben werden. Dies ermöglicht eine einfache Übersicht, welche Themen und Teilfächer aktuell – und bei wiederholten Auswertungen im Zeitverlauf – von besonderem Gewicht sind.

Für die Ermittlung der Akzeptanz von Medien sollen die Beurteilungen der Medien analysiert werden. Neben der Auswertung der numerischen Benotung der Medien von 5 (höchster Wert) bis 1 (niedrigster Wert) sollen auch die im Freitextfeld ein-

gegebenen Kommentare in die Analyse einbezogen werden. Das konkrete Analyseverfahren hängt von den vorhandenen Textdaten ab. Diese müssen zunächst linguistisch aufbereitet werden. Vorstellbar wäre beispielsweise, Häufigkeitslisten der in den Rezensionen verwendeten Wörter zu erstellen, auf deren Grundlage dann intellektuell die Terme extrahiert werden, denen im Hinblick auf Wertungen und unter Berücksichtigung von möglicherweise relevanten Themengebieten ein besonderes Gewicht zukommt. Durch die Zuordnung von Synonymen zu diesen Termen kann die Menge der berücksichtigten Wörter aus den Rezensionen erweitert werden. Die auf diese Weise strukturierten Textdaten können dann mit Methoden des Data Mining weiter analysiert werden. Möglich wäre es, die Rezensionen vorgegeben Kategorien (z. B. nach Fach und Bewertung) zuzuordnen, sie zu ähnlichen Gruppen zusammenzufassen oder das gemeinsame Auftreten von Termen zu analysieren (vgl. Hippner/Rentzmann 2006). Darüber hinaus können diese aufbereiteten Daten im Kontext mit weiteren Variablen, wie z. B. der Anzahl der Vormerkungen, häufig verwendeten Schlagwörtern und RVK-Notationen untersucht werden.

#### **4.5 Steuerung der betrieblichen Abläufe**

Die Berücksichtigung der Rezensionseinträge und der durch die statistischen Verfahren gewonnenen Erkenntnisse ermöglicht ein deutlicher auf die Zielgruppen ausgerichtetes Angebot von Produkten und Dienstleistungen von Hochschulbibliotheken als bisher. Durch die Auswertung der Rezensionen kann die Bibliotheksleitung erkennen, welche Defizite bei den angebotenen Medien bestehen. Sie erhält zusätzliche Anhaltspunkte darüber, welche Bestände stark und wenig nachgefragt werden und vor allem wie diese Bestände bewertet werden. Auf der Grundlage dieser Informationen kann überlegt werden, ob und wie bestimmte Bestände (z. B. die elektronischen Datenbanken) anders präsentiert und beworben werden müssen bzw. in welchen Bereichen die Bestände erweitert werden sollten.

Im Erwerbungsbereich können durch die gewonnenen Erkenntnisse Schwerpunkte gesetzt werden. Beispielsweise ist es sinnvoll, weitere Lehrbücher zu bestellen, wenn bestimmte Titel mittels Rezensionen empfohlen werden. Gerade diese zusätzlichen Informationen fehlen häufig bei der Steuerung der Erwerbung. Aus der Anzahl der Vormerkungen eines Werkes kann nur bedingt auf die Bedeutung für die Lehre geschlossen werden, da viele Studierende Medien wahllos vormerken bzw. bei zu langen Wartezeiten vollständig auf eine Vormerkung verzichten.

Auf diese Weise können nicht nur einzelne Titel, sondern auch Themengebiete ermittelt werden, in denen der verstärkte Kauf von Medien sinnvoll ist. Beispielsweise kann die Bibliotheksleitung veranlassen, dass 10 % des Erwerbungsetats der

Betriebswirtschaftslehre für das Thema Marketing eingesetzt werden. Es geht in diesem Zusammenhang aber auch darum, die richtige Relation zwischen der Berücksichtigung der Kundeninteressen und einem langfristigen möglichst breit angelegten Bestandsaufbau zu bestimmen.

Durch den Einsatz von Online-Marketing wird in der Privatwirtschaft die Wertschöpfungskette auf den Kopf gestellt. In letzter Konsequenz wird ermöglicht, dass primär der Kunde das Leistungsniveau bestimmt und erst an zweiter Stelle der Anbieter darauf reagiert. Eine Übertragung auf Hochschulbibliotheken führt zur konsequenten Umsetzung des Konzepts der Bringbibliothek und damit zur Steuerung der Betriebsabläufe unter Berücksichtigung der Kundeninteressen. Die Bibliothek ist dadurch in der Lage, flexibel und zeitnah auf die Wünsche der Kunden zu reagieren.

Insgesamt sollen die Ergebnisse der Evaluation zu einer besseren Steuerung von wichtigen Geschäftsprozessen in der Bibliothek beitragen. Dabei stehen vor allem eine stärker an den Kundenbedürfnissen ausgerichtete Erwerbung, ein für die Kunden attraktives Schulungsangebot und der optimale Einsatz der Mitarbeiterinnen und Mitarbeiter im Vordergrund.

## **5 Fazit und Ausblick**

Bibliotheken müssen die Möglichkeiten der Informationstechnologien und innovativer Marketingstrategien nutzen und ihr Angebot konsequent auf die Bedürfnisse ihrer Kunden ausrichten, wenn sie im Wettbewerb mit anderen Informationsanbietern erfolgreich bestehen wollen. Um die eigenen Geschäftsprozesse im Hinblick auf eine stärkere Kundenorientierung neu zu konzipieren, ist es notwendig, aktiv den Dialog mit den Kunden zu suchen und Informationen über sie und ihre konkreten Wünsche zu gewinnen. Dafür bieten sich im Rahmen von Web 2.0 Anwendungen zahlreiche Möglichkeiten, die derzeit auch an anderen Hochschulbibliotheken ausprobiert werden. Neben Weblogs, die im Bereich der Kommunikation nach außen eingesetzt werden, wird es in Zukunft noch stärker darum gehen, die Kunden interaktiv und kooperativ an der konkreten Ausgestaltung der Angebote zu beteiligen. Dabei sind neben dem Erwerbungsbereich auch Modelle für eine kollaborative Erschließung von Medien denkbar, bei denen Kunden – ähnlich dem Tagging bei Web 2.0 Bookmarking Diensten wie *delicio.us* oder *Connotea* – Medien der Bibliothek mit eigenen Schlagwörtern kennzeichnen. Noch sind die Reaktionen auf diese Angebote eher verhalten. In dem Maße wie sich das Internet weiter wandelt und die Kunden diese Kommunikationsformen verinnerlichen, werden sie

den Anspruch des Mitgestaltens auch in Hochschulbibliotheken und ähnliche Einrichtungen hineinragen. Nur in enger Kooperation mit den Kunden kann es Hochschulbibliotheken gelingen, den kommenden Erwartungen an Benutzerfreundlichkeit, Interaktivität und Flexibilität gerecht zu werden.

## 6 Literatur

- blogscout.de (2007): Die Top-Reichweite-Blogs vom 09.04.2007, [http://blogscout.de/blogstop\\_reichweite\\_taeglich](http://blogscout.de/blogstop_reichweite_taeglich), Zugriff 10.04.2007.
- Conrady, Roland et al. (Hrsg.) (2002): Online-Marketing-Strategien: Konzeption, Technologien, Prozesse, Recht. Neuwied u a.: Luchterhand.
- Halle, Axel (2005): Chancen und Risiken der Bibliotheken im Informationszeitalter: vom Knowbody zum Nobody? In: Kolding-Nielsen, Erland et al. (Hrsg.): Die innovative Bibliothek. München: Saur, S. 29-39.
- Hippner, Hajo/Rentzmann, Rene (2006): Text Mining zur Anreicherung von Kundenprofilen in der Bankbranche. In: HMD – Praxis der Wirtschaftsinformatik, Heft 249, S. 99-108.
- Huber, Andreas (2006): Marketing. 2., vollst. überarb. und erw. Aufl., München: Vahlen.
- Kleinwächter, Wolfgang (2007): Die nächste Milliarde Menschen online bringen. In: Telepolis, <http://www.heise.de/tp/r4/artikel/24/24797/1.html>, Zugriff 08.03.2007.
- Lange, Corina (2006): Web 2.0 zum Mitmachen. Die beliebtesten Anwendungen. Köln: O'Reilly.
- O'Reilly, Tim (2005): What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software, <http://www.oreilly.de/artikel/web20.html> Zugriff 20.01.2007.
- Übersetz. von Patrick Holz (2006): Was ist Web 2.0? Designvorlagen und Geschäftsmodelle für eine neue Software-Generation, [http://twozero.uni-koeln.de/content/e14/index\\_ges.html](http://twozero.uni-koeln.de/content/e14/index_ges.html), Zugriff 20.01.2007.
- Schmidt, Jan (2006): Weblogs: Eine kommunikationssoziologische Studie. Konstanz: UVK.
- Schroll, Willi/Rodenhäuser, Ben/Neef, Andreas (2007): Mash Up Your Business! Der Web 2.0 Report. Essen u a.: Z\_punkt GmbH The Foresight Company, <http://web2.oreport.de/public/Leseprobe.pdf>, Zugriff 20.01.2007.
- Schwarz, Peter (2001): Management-Brevier für Non-Profit-Organisationen: eine Einführung in die besonderen Probleme und Techniken des Managements von privaten Non-Profit-Organisationen. 2., vollst. überarb. und erw. Aufl., Bern u a.: Haupt.
- Universitätsbibliothek Mannheim (2007) Rezensionenrichtlinien, <http://mals.bib.uni-mannheim.de/mopa/Rezensionsrichtlinien.html>, Zugriff 06.03.2007.
- Weigend, Andreas (2006): Ein riesiges Verhaltenslabor. Interview von Philipp Wolff. In: SZ Wissen 13/2006, <http://www.sueddeutsche.de/wissen/artikel/983/94889/>, Zugriff 20.01.2007.

WIR (Wissenschaftsrat) (2001): Empfehlungen zur digitalen Informationsversorgung durch Hochschulbibliotheken, <http://www.wissenschaftsrat.de/texte/4935-01.pdf>, Zugriff 20.01.2007.

youtube.com (2007): Most Viewed (All Time)

<http://www.youtube.com/browse?s=mp&t=a&c=0&l=>, Zugriff 10.04.2007.

Zerfass, Ansgar (2005): Corporate Blogs: Einsatzmöglichkeiten und Herausforderungen, <http://www.zerfass.de/CorporateBlogs-AZ-270105.pdf>, Zugriff 20.01.2007.

Zerfass, Ansgar (2006): Die neuen Meinungsmacher. Bedeutung, Chancen und Risiken für Unternehmen. Vortrag auf der 5. Internationalen Konferenz der Issues Management Gesellschaft Deutschland e. V. am 08.09.2006, Wiesbaden, [http://www.image-ev.com/downloads/Konferenz\\_2006/Zerfa%DF\\_Meinungsmacher.pdf](http://www.image-ev.com/downloads/Konferenz_2006/Zerfa%DF_Meinungsmacher.pdf), Zugriff 10.04.2007.



# Von der Digitalisierung der Objekte zur Virtualisierung der Kulturgüter

## Stand des Innovationsprozesses am Beispiel der Museen der Schweiz\*

*Christina Bieber*

Swiss Institute for Information Research  
HTW Chur  
Ringstrasse/ Pulvermühlestrasse 57  
CH-7004 Chur  
*Christina.Bieber@fh-htwchur.ch*

### **Zusammenfassung**

Nach einem Überblick zu der bisherigen Entwicklung digitaler Zugriffsmöglichkeiten auf Kulturgüter aus dem Bereich von Bibliotheken und Archiven werden Formen und Arten der Digitalisierung von Museumsobjekten beschrieben. Es wird der Prozess der zunehmenden Virtualisierung von der Erstellung der digitalen Basisdaten über verschieden elaborierte Web-Präsenzen bis zu umfassenden kooperativ erstellten Informations- und Dienstleistungsangeboten aufgezeigt. Im Anschluss wird der Entwicklungsstand für die Museen der Schweiz beschrieben. Grundlage hierfür ist eine am Schweizerischen Institut für Informationswissenschaft der HTW Chur erstellte Studie.

### **Abstract**

The first part of this paper provides an overview of recent developments in accessing cultural heritage information on the web. A short introduction to documentation and digitization in the field of museums will be given. The next part describes the process of digitizing and virtualizing museum artefacts. A categorization is presented illustrating the development from simple web sites, their shift to more elaborate ones and their latest transformation into service portals making use of collaborative elements and integrating different sources of content. In the last part the

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 221-236.

status quo of the digitization efforts in museums in Switzerland is depicted. Data is based on a study recently carried out by the Swiss Institute for Information Research.

## I Digitalisierung von Kulturgütern – ein kurzer Überblick

Museen, Bibliotheken und Archive gehören zu den Einrichtungen, die eine Vielzahl von Gütern beherbergen, die wesentliche Teile unserer kulturellen Identität repräsentieren. Im englischsprachigen Bereich hat sich für diese Art Einrichtungen der übergreifende Begriff der *Memory Institutions* herausgebildet. Aus diesem Dreierbund waren es in Deutschland – ähnlich der Entwicklung in vielen anderen Ländern – die Bibliotheken, die als erste ihre Bestände umfangreich im Netz recherchierbar machten. Mittlerweile werden den Online-Kunden vielfältige Dienste und Funktionen rund um die gewünschten Informationen zur Verfügung gestellt. Zu den jüngsten Angeboten gehören zum Beispiel Volltext-Digitalisate einzelner Artikel oder kompletter Werke, bei denen das Urheberrecht erloschen bzw. die Rechtslage geklärt ist. Digitale Publikationen haben Aufnahme in die nationale Verzeichnung, die elektronische Aufbewahrung und das allgemeine Medienangebot gefunden. Katalog-Daten werden u. a. mit Inhaltsverzeichnissen angereichert und erlauben so eine gezielte Auswahl der Literatur<sup>1</sup>. Portal-Lösungen ermöglichen komfortablere Meta-Recherchen in zuvor einzeln zu konsultierenden Datenbeständen. Personalisierungsmöglichkeiten wie etwa MyLibrary-Konzepte finden zunehmend Anwendung und einige Bibliotheken bieten einen virtuellen Auskunftsdienst<sup>2</sup> an. Auf europäischer Ebene wird die Entwicklung der European Digital Library<sup>3</sup> mit multilingualen Zugriffsoptionen voran gebracht [Niggemann 2006]. Dem traditionsreichen, national und international gut vernetzten Bibliotheksbereich mit seinem vergleichsweise homogenen Sammlungsgut, das nach nationalen und international kompatiblen Standards erfasst wird, war und ist es relativ schnell möglich, die Funktionalitäten der Informationstechnologie und des Webs zu nutzen.

---

<sup>1</sup> Datenbank SWBplus des Südwestdeutschen Bibliotheksverbundes enthält ergänzende Informationen wie z. B. Inhaltsverzeichnisse, Rezensionen, Klappentexte, Verlagsinformationen und Abstracts [http://www.bsz-bw.de/SWBplus/sp\\_recherche.shtml](http://www.bsz-bw.de/SWBplus/sp_recherche.shtml) Stand 03.01.2007.

<sup>2</sup> Online-Auskunft per Chat der Universitäts- und Landesbibliothek Düsseldorf <http://www.ub.uni-duesseldorf.de/home/service/onlineauskunft> Stand 03.01.2007.

<sup>3</sup> European Digital Library, Projektlaufzeit: 1.09.2006 – 29.02.2008. Projekthomepage: <http://www.edlproject.eu> Stand 03.01.2007.

Etwas differenzierter ist die Entwicklung bei den Archiven zu betrachten. Bedingt durch verschiedene Träger, Funktionen und Medientypen gestaltet sich die Archivlandschaft vielfältiger und die Entwicklung ist weniger einheitlich. Die Findmittel liegen in den größeren Einrichtungen im deutschsprachigen Raum zunehmend digital vor. Die Landesarchive und vergleichbare Einrichtungen mit Pflichtlegat stehen vor der Herausforderung, für die im Zuge von eGovernment entstehenden elektronischen Akten und Dokumente ganzheitliche Konzepte zu entwickeln [Rehm 2006].

In den Medienarchiven gehören die Audio- und Filmmaterialien zu den gefährdeten Kulturgütern unserer Zeit. Aktuell existieren verschiedene weltweite Initiativen, nationale Programme und zahlreiche Einzelprojekte, die sich der langfristigen Sicherung und Digitalisierung dieser Quellen widmen. Die Grundlage vieler dieser Aktivitäten bildet die internationale Kultur- und Bildungsorganisation UNESCO (United Nations Educational, Scientific and Cultural Organization). "Preservation towards Storage and Access, Standardized Practices for Audiovisual Contents in Europe" hat das EU-Projekt PrestoSpace zum Inhalt.<sup>4</sup> Als großes Einzelprojekt sei das der BBC angeführt, die ihren Katalog mit rund 1 Million Rundfunk- und Fernsehprogrammen in einem Prototyp online gestellt hat.<sup>5</sup> Das Ziel des EU-finanzierten Video Active-Projektes ist die Schaffung eines gemeinsamen Zugangsportals zu den Fernseharchiven in Europa.<sup>6</sup>

Am Beispiel der Bibliotheken und Archive zeigt sich deutlich, dass nachdem der elektronische Zugriff auf die Basisdaten der Bestände realisiert wurde, weiterführende Dienste und Funktionen angeboten sowie Kooperationen eingegangen werden. Diese Kooperationen können dem Ziel folgen, eine geographisch größere, gegebenenfalls übernationale und mehrsprachige Abdeckung zu erreichen oder das Spektrum der angebotenen Informationen wie auch Dienstleistungen zu erweitern. Ein deutsches Beispiel für letzteres ist das in Baden-Württemberg angesiedelte gemeinsame Portal von Archiven, Bibliotheken und Museen [Maier 2005]. Größer angelegt ist das Scottish Cultural Resources Access Network (SCRAN)<sup>7</sup>. Das als Millennium-Projekt im Jahre 2000 von der britischen Regierung finanzierte Portal

---

<sup>4</sup> PrestoSpace: Start im Februar 2004 im einer Laufzeit von 40 Monaten <http://www.prestospace.org/> Stand: 03.01.2007.

<sup>5</sup> BBS Program Catalogue <http://open.bbc.co.uk/catalogue/infax/> Stand: 03.01.2006.

<sup>6</sup> Video Active <http://videoactive.wordpress.com/> Stand: 04.01.2007.

<sup>7</sup> SCRAN <http://www.scran.ac.uk> Stand: 04.01.2007.

ermöglicht aktuell den Zugriff auf über 1 Million Medien aus ca. 400 Museen, Archiven und Bibliotheken Schottlands.

Die Aufgaben von Museen werden allgemein mit Sammeln, Bewahren und Vermitteln definiert [ICOM 1990]. Im Bereich der Vermittlung über das Internet bestehen für Museen noch große Entwicklungspotenziale. Die Wahrnehmung dieser Aufgaben kann durch Informationstechnologien wesentlich unterstützt werden. In diesem Beitrag liegt der Schwerpunkt bei Technologien, Anwendungen sowie Dienstleistungen über das Internet. Die hierbei üblichen technischen Verfahren, die klassische Objektdokumentation sowie die neueren Möglichkeiten sollen im Folgenden skizziert werden. Anschließend wird in quantitativer Hinsicht der Stand des Innovationsprozesses am Beispiel der Museen der Schweiz beschrieben.

## **2 Von der Digitalisierung der Objekte zur Virtualisierung der Kontexte**

### **2.1 Erstellung der digitalen Basisdaten im Museumsbereich**

Mit Digitalisierung wird allgemein die Umwandlung von analogen Informationen in eine digitale Form bezeichnet. Im Museumsbereich bestehen kurz gefasst mehrere Anwendungsfelder für eine Digitalisierung:

- das Museumsobjekt selbst
- Informationen zum Objekt
- weitere Materialien mit Kontextbezug.

An erster Stelle können die Museumsobjekte selbst digital repräsentiert werden. Hier ist an die visuelle Repräsentation gedacht. Dies geschieht heute häufig in Form der Digitalfotografie. Scanner werden üblicherweise etwa im Bereich der Fotografie bzw. für Filmnegative eingesetzt. 3D-Scans sind für dreidimensionale Objekte möglich. Die 3D-Technologien finden jedoch eher im Bereich der Rekonstruktion und der Animation Anwendung [vgl. Hennings et al. 1996]. Für den Bereich der audiovisuellen Medien (Tonträger, Fernsehen und Rundfunk) stehen heute verschiedene effiziente Digitalisierungstechnologien zur Verfügung. Viele nationale und internationale Initiativen und Projekte bemühen sich gerade um die Digitalisierung großer audiovisueller Bestände.

An zweiter Stelle für die Digitalisierung stehen die vorhandenen Informationen zum Objekt. Viele Objekte, die zur Zeit ihrer Entstehung selbsterklärend waren, verloren diese Eigenschaft für die nachkommenden Generationen. Die Bedeutung

eines Kulturgutes ist daher eng verbunden mit dem Vorhandsein zugehöriger Informationen zu seiner Entstehung, gegebenenfalls zur Verwendung oder historischen Bedeutung. Eine Objektbeschreibung beinhaltet üblicherweise die Provenienz, Gattung, Materialerfassung, Maße, Erhaltungszustand, Abbildung, Literaturangaben oder Quellen sowie Verweise auf ähnliche oder zugehörige Objekte.

An dritter Stelle stehen weitere Materialien vielfältiger Ausprägung, die weniger zu einzelnen Objekten als zu größeren Einheiten gehören und meist ein Kontextverständnis ermöglichen.

Daraus resultieren drei Hauptformen von digitalen Daten:

- digitale Originalobjekte
- Digitalisate
- digitale Sammlungsdokumentation.

Digitale Originalobjekte bezeichnen originäre digitale Daten, die nicht durch die Umwandlung von analogen in digitale Informationen entstanden. Hierunter fallen etwa Digitalfotografien, 2D- oder 3D-Rekonstruktionen sowie Animationen oder beispielsweise Netzkunst. Aber auch fast alle Texte, die im Museum entstehen, sei es für die Wandbeschriftung, den Ausstellungskatalog oder pädagogische Materialien für den Download im Web fallen darunter.

Digitalisate sind Objekte, die durch Umwandlung analoger Signale in die digitale Form entstanden sind. Meist handelt es sich um Schrift-, Text- oder Bildzeugnisse. Beispielsweise sind hier gescannte Fotografien oder Zeitungen zu nennen.

Die digitale Dokumentation bezeichnet die digitale Verzeichnung der Sammlungsobjekte. Diese sollte konsistent sein und auf Standards beruhen. In der Basis erfasst sie die am Objekt selbst ablesbaren Informationen (z. B. Maße, Material, Alter), Angaben zur Herkunft oder Funktion sowie weitere Merkmale. Unter Umständen kann es ein Instrument mit wissenschaftlichem Anspruch sein, das den gegenwärtigen Forschungsstand aufzeigt. Die auf dem Markt verfügbaren Softwareprodukte haben in den vergangenen Jahre Funktionen zur Prozessunterstützung rund um das Objekt integriert. Man spricht daher auch häufig von einer digitalen Objektverwaltung. Die klassische Objektdokumentation liefert dabei nach wie vor die Grundlage für alle weiteren Vorgänge rund um das Objekt. Dies sind im Einzelnen die allgemeine interne oder externe wissenschaftliche Kommunikation, die Ausstellungstätigkeit, die Öffentlichkeitsarbeit, die Museumspädagogik, sowie das gesamte Publikationswesen [Waidacher 1999].

Eine medienneutrale Datenhaltung, das Crossmedia-Publishing und damit Anforderungen an ein professionelles Informationsmanagement haben somit auch in den Museen Einzug gehalten.

## **2.2 Verschiedene Stufen der Virtualisierung**

Mit Virtualisierung sollen hier der fortschreitende Prozess des Ausbaus und der Weiterentwicklung der Konzepte für eine Web-Präsenz bezeichnet sein. Ausgehend von einer explorativen Studie nahm Wersig und Schuck-Wersig [Wersig et al. 1997] folgende formale, überwiegend auf den Umfang bezogene Kategorisierung vor:

- **Minimalinformatives Angebot**  
Bestehend aus den Angaben Name, Anschrift, Öffnungszeiten u ä. in der Grundform. In der erweiterten Form werden Zusatzinformationen geliefert mit Abbildungen, Hinweisen etwa zum Ausstellungs- und Veranstaltungskalender, zur Sammlung und zur Geschichte des Hauses.
- **Erweitertes Informationsangebot**  
Gegenüber der Grundform bietet dieses mehr Informationen und ermöglicht so eine umfassendere und flexiblere Nutzung. Dies ist meist mit höherem technischem Aufwand verbunden.
- **Umfangreiches Informationsangebot mit zahlreichen Links**  
Hohes Niveau von Technik und Design, umfangreiche Informationen und vielfältige Optionen.

Diese Einteilung war vor allem für die Pionierjahre der Museen im Netz hilfreich. Mehr auf die Inhalte bezogen ist die Einteilung die Schweibenz [Schweibenz 1998] erarbeitete:

- **Broschürenmuseum**  
Dieses Angebot hat informierenden und werbenden Charakter. Es ist vergleichbar mit dem minimalinformativen bzw. erweiterten Informationsangebot nach Wersig / Schuck-Wersig.
- **Inhaltsmuseum**  
Es beinhaltet alle zuvor beschriebenen Komponenten und ermöglicht den Zugang zur Objektdatenbank. Dies jedoch ohne didaktische Aufbereitung und ist daher eher für ein Fachpublikum geeignet.
- **Lernmuseum**  
Beim Lernmuseum handelt es sich um eine Weiterentwicklung des Inhaltsmuseums.

Es bietet didaktisch aufbereitete Angebote für verschiedene Zielgruppen. Statt einer Objektorientierung steht hier eine Kontextorientierung im Mittelpunkt.

Die Internetauftritte kleinerer Museen in der Schweiz sind heute mehrheitlich dem Typus des Broschürenmuseums zuzurechnen. Die meisten größeren Einrichtungen im deutschsprachigen Bereich befinden sich in der Entwicklung vom Inhalts- zum Lernmuseum. Weiter vorangeschritten ist diese Entwicklung beispielsweise in den englischsprachigen Ländern [vgl. auch Bieber et al. 2007].

Mit Rückbezug auf die zuvor angesprochenen jüngeren Kooperationsformen können die folgenden Kategorien ergänzt werden:

- Meta-Museum  
Die Objekte mehrerer Museen sind gemeinsam recherchierbar.
- Meta-Portal für Kulturgüter  
Die Bestände einer Reihe von unterschiedlichen Einrichtungen (z. B. Museen, Archive, Bibliotheken, Forschungseinrichtungen) können konsultiert werden.
- Meta-Projekte unterschiedlichen Charakters wie z. B. kooperativ erstellte virtuelle Ausstellungen.

Die angeführten Aspekte zeigen den beschrittenen Weg der letzten 10-15 Jahre auf, der von einer ersten Informatisierung über die Digitalisierung der Objekte und einer internen Prozessunterstützung zur Kontextualisierung der Vermittlung ging.

Seit etwa 2002 etablieren sich im Internet neue technische Features und Community-Phänomene, die allgemein unter dem Begriff Web 2.0 subsumiert werden [O'Reilly 2005]. Hier bestehen für Museen Potenziale in der Präsentation und Interaktion. Bislang sind Beispiele im deutschsprachigen Bereich jedoch rar. Im Hinblick auf die neuere Technologien könnten RSS-Feeds ohne Zweifel ein guter Informationskanal sein. Es ist davon auszugehen, dass Podcasts zunehmend die bisherigen Formen des Datenstreamings ablösen. Audioguides für die Führung vor Ort werden sehr wahrscheinlich zukünftig durch Podcasts auf ein persönliches Consumer Device ersetzt werden. Es ist zu fragen, inwiefern Aspekte im Bereich der Social Software oder Personalisierung Akzeptanz finden, da hier letztlich das Selbstverständnis von Museen berührt wird. Die Konzepte von Social Software sehen den Nutzer des Webs nicht nur als Konsumenten, sondern als einen Akteur, der im Austausch mit Anderen steht. Diese „aktive“ Rolle ist in der deutschsprachigen Museumskultur allenfalls im Rahmen von museumspädagogischen Programmen vorgesehen. MyGallery-Konzepte finden sich im englischsprachigen Bereich bereits seit Jahren. Beispiels-

weise bieten das Metropolitan Museum of Art<sup>8</sup> oder das Fine Arts Museum of San Francisco<sup>9</sup> ihren virtuellen Besuchern die Möglichkeit, Werke auszuwählen und in einem eigenen Bereich zu arrangieren. Im deutschsprachigen Bereich haben diese Beispiele bisher keine Akzeptanz gefunden.

### **3 Stand der Digitalisierung in den Museen der Schweiz – Eine Studie zum Thema**

Im Auftrag des Verbandes der Museen der Schweiz (VMS) wurde am Schweizerischen Institut für Informationswissenschaft der HTW Chur 2006 eine Studie zum Stand der Digitalisierung im Museumsbereich der Schweiz durchgeführt.

Ziel der Studie war es, gesicherte Aussagen über den Stand der Digitalisierung, Art und Formen digitaler Angebote im Web, Informationen zu den Zielsetzungen und Prioritäten bei der Digitalisierung, sowie Problembereiche und Lösungswünsche zu erfassen. Hierzu wurden eine Online-Umfrage sowie eine umfangreiche Web-Recherche durchgeführt. Das Bild ergänzen internationale Referenzprojekte. Zusammenfassend konnten für die Schweiz Handlungsempfehlungen formuliert werden. Die nachfolgenden Ausführungen beruhen auf statistischen Teilergebnissen dieser Studie. Für Good Practice-Beispiele sowie Detailbetrachtungen muss an dieser Stelle auf die Studie selbst verwiesen werden [Bieber et al. 2007].

#### **3.1 Online-Umfrage**

Im Sommer 2006 umfasste das Museumsverzeichnis des Verbandes der Museen der Schweiz insgesamt 1.019 Museen. Zu 819 Einrichtungen konnten die für eine Online-Umfrage notwendigen Email-Adressen ermittelt werden. Die Umfrage, deren Fragebögen je nach Adressaten in einer der drei Landessprachen verschickt wurden, erfolgte webgestützt mittels des Befragungstools 2ask. Der Fragebogen enthielt 14 Hauptfragen und zahlreiche Unterkategorien. In diesem Beitrag sollen eine Auswahl der Ergebnisse präsentiert werden.

Mit einer Rücklaufquote von 32 % – dies entspricht 249 Museen – ist die Beteiligung als gut zu bewerten. Die folgende Abbildung zeigt die Verteilung der teilnehmenden Einrichtungen auf verschiedene Museumstypen. Mehrfachnennungen waren möglich.

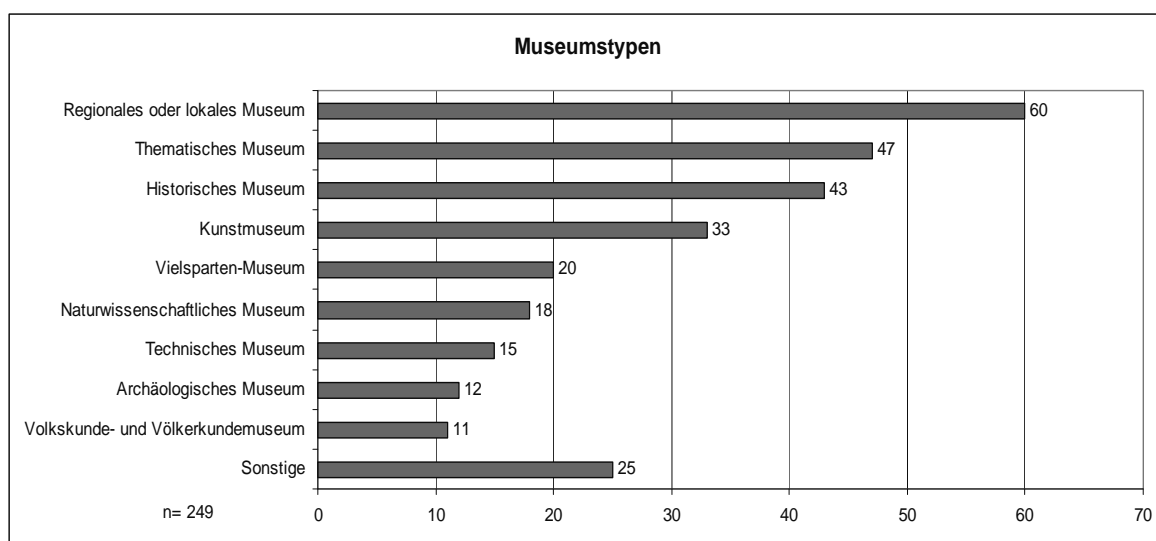
---

<sup>8</sup> Metropolitan Museum of Art <http://www.metmuseum.org> Stand: 04.01.2007.

<sup>9</sup> Fine Arts Museum of San Francisco <http://www.thinker.org> Stand: 04.01.2007.



Hierzu wurde ein Vergleich mit der offiziellen Verbandsstatistik aus dem Jahre 2003 durchgeführt. Die Kategorisierung ist in Teilen unterschiedlich. Es ließ sich jedoch schlussfolgern, dass hinsichtlich der Art sowie Verteilung der teilnehmenden Museen deutliche Parallelen und eher geringere Abweichungen festgestellt werden können und daher von repräsentativen Aussagen der Studie auszugehen ist. Es liegen ferner von 231 Häusern Angaben zu den Stellenprozenten bzw. zu einer ehrenamtlichen Führung vor. Von diesen werden rund 28 % ehrenamtlich geführt; 29 % verfügen über maximal eine Stelle. Zusammengenommen entfallen damit ca. 57 % auf Kleinstmuseen. Bis zu 10 Stellen finden sich bei rund 34 % der hier antwortenden Häuser. Grosse Häuser mit über 10 Stellen machen etwa 8 % aus.



*Abbildung 1: Verteilung nach Museumstypen*

Digitale Daten liegen bei 71 % der antwortenden Häuser vor, während 29 % noch keine Digitalisierungen vorgenommen haben. Rund 75 % der Häuser ohne bisherige Digitalisierung planen auch für die nächste Zukunft keine Aktivitäten in diesem Bereich, während 25 % sehr wohl Planungen haben. Damit würde sich in nächster Zeit der Anteil der Häuser mit Digitalisaten von 71 % auf rund 78 % erhöhen.

Als Gründe für eine bisher nicht erfolgte Digitalisierung nannten 43 % finanzielle Aspekte. Fehlende Konzepte nannten 39 % der Museen.

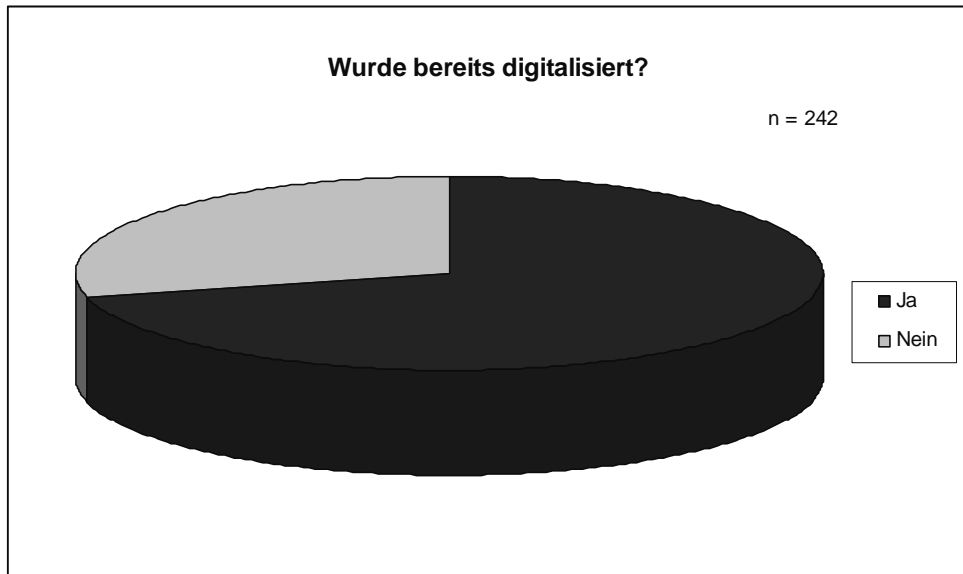


Abbildung 2: Vorhandensein von Digitalisaten

Die nachfolgende Graphik zeigt die allfälligen digitalen Angebote der 249 antwortenden Museen. Am häufigsten ist eine digitale Objektverzeichnung vorhanden. Bei dieser Frage wurde nicht zwischen intern vorgehaltenen und extern über das Internet zugänglichen Verzeichnissen unterschieden. Als nächstes folgen Digitalisate zu einzelnen Exponaten sowie verschiedene Materialien. Virtuelle Museen, virtuelle Ausstellungen oder Metadatenbanken sind in Einzelbeispielen vertreten.

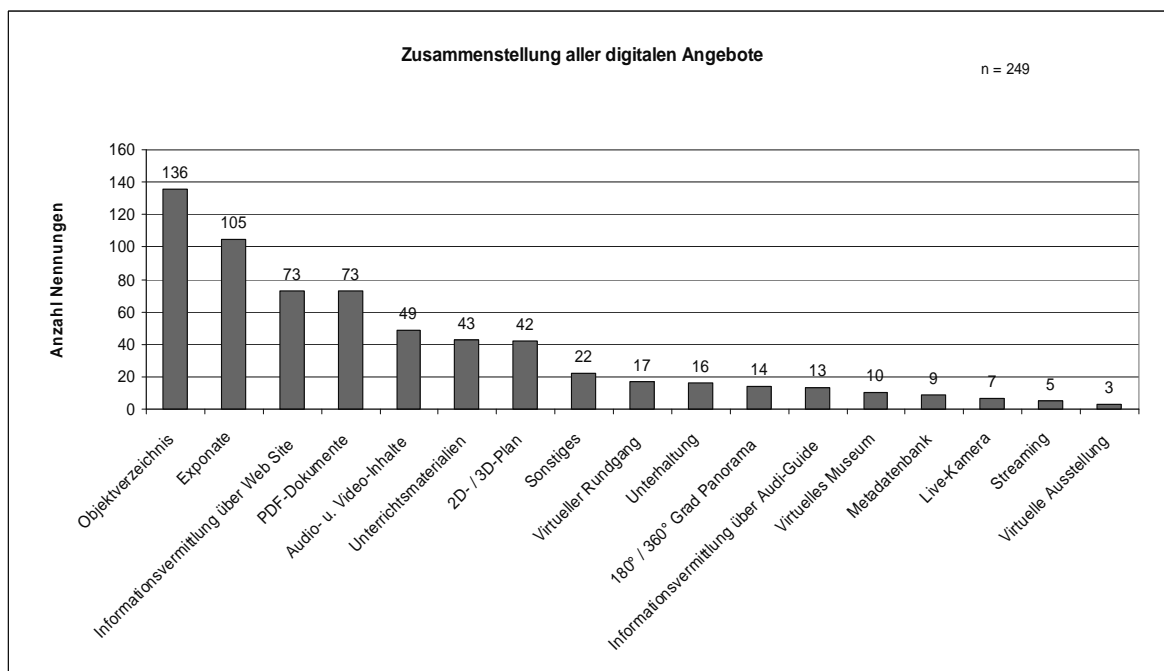


Abbildung 3: Angebotsformen

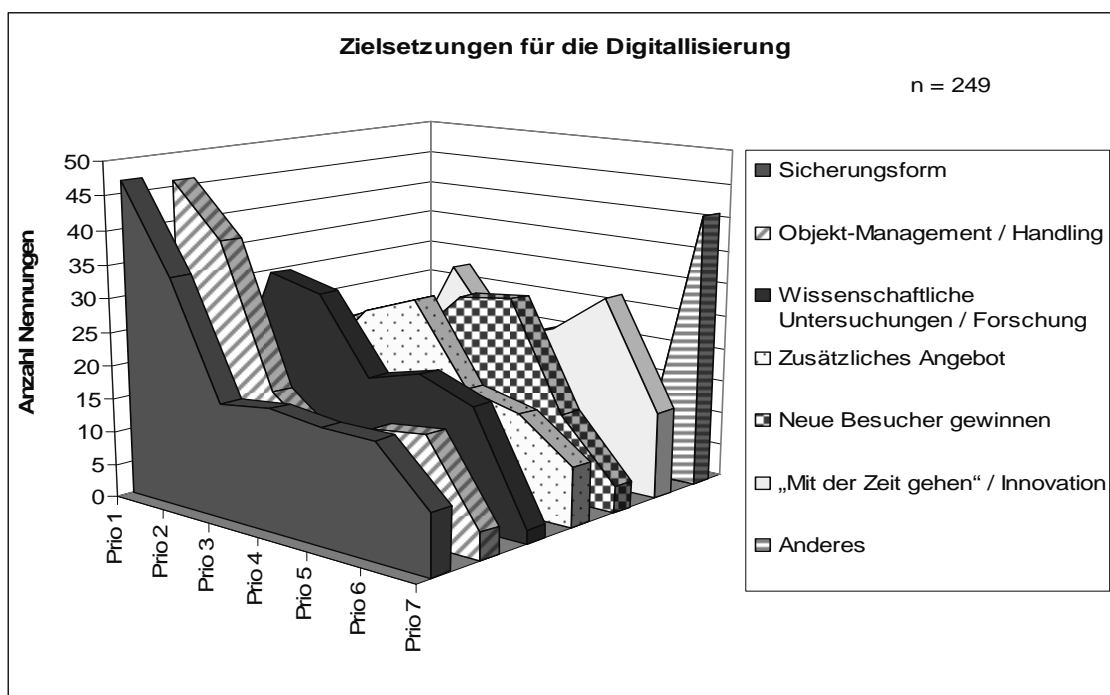


Abbildung 4: Prioritäten bei der Digitalisierung

Gefragt nach den verfolgten Zielsetzungen und der Priorisierung dieser auf einer Skala von 1 (höchste Priorität) bis 7 (niedrigste Priorität) steht die Digitalisierung als Form der Sicherung an erster Stelle. An zweiter Stelle steht die Absicht, das Objekthandling zu unterstützen.

Die nachfolgende Graphik verdeutlicht, dass je mehr Mitarbeiter vorhanden sind, desto wahrscheinlicher erfolgt die Objektverzeichnung digital.

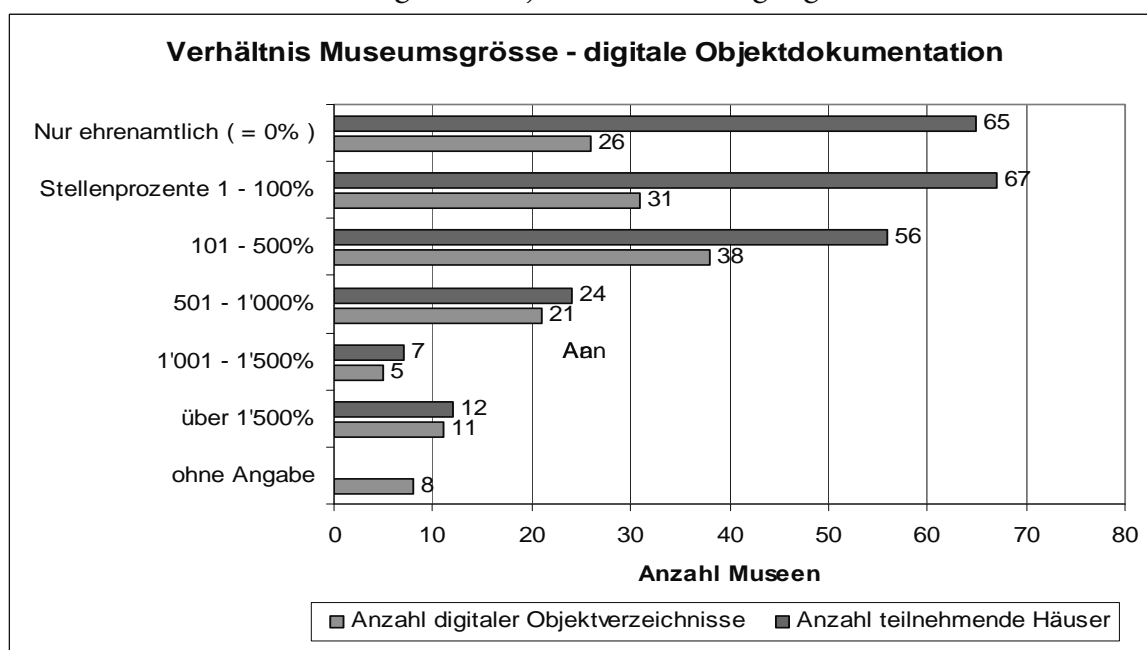


Abbildung 5: Verhältnis Museumsgröße und digitale Objektdokumentation

Nur etwa 26 % der Museen verwenden für die Objektverzeichnung vordefinierte Standards. Etwa die Hälfte nutzt selbst entwickelte Schemata.

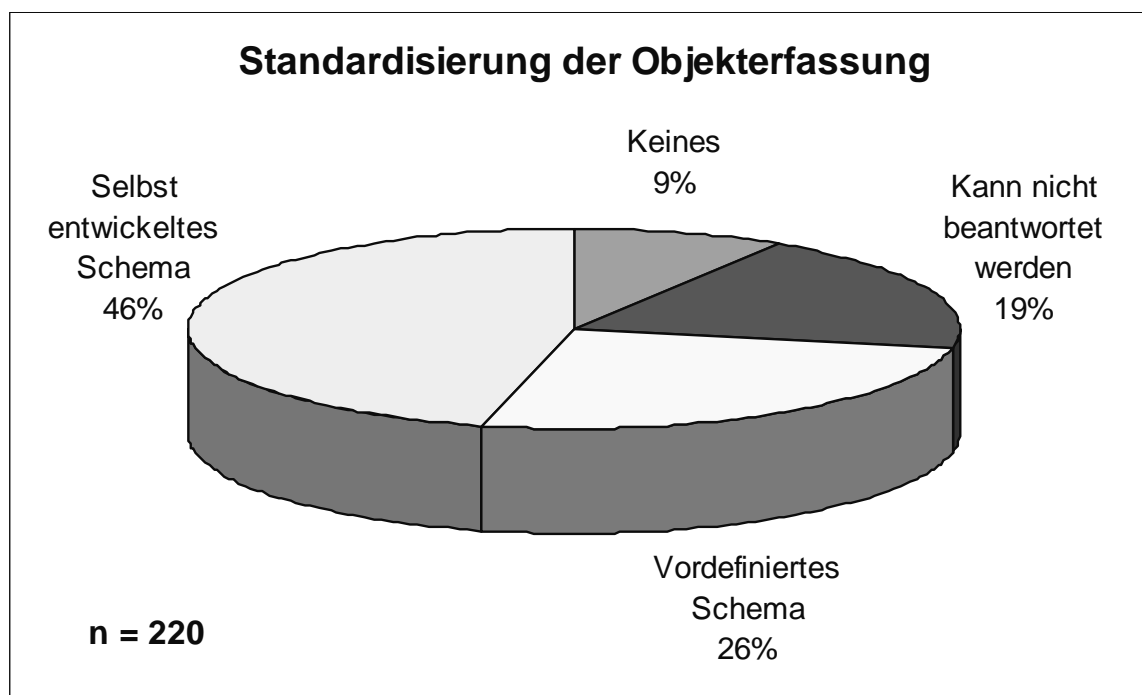


Abbildung 6: Standardisierung der Objekterfassung

### 3.2 Web-Recherche

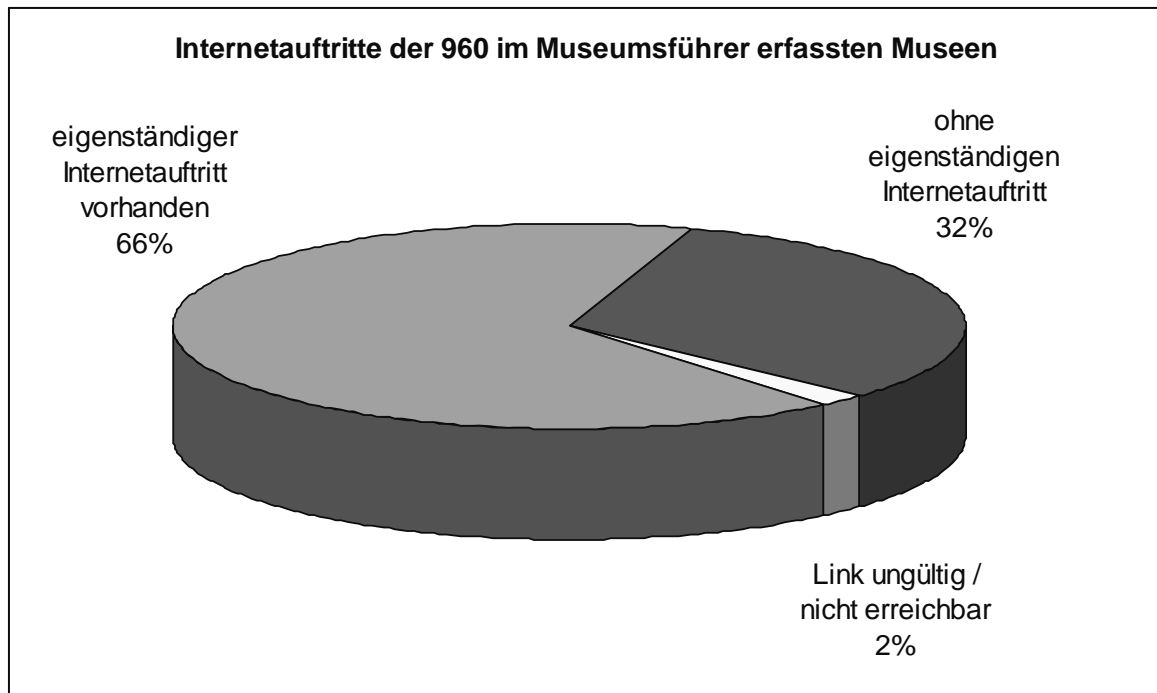
Im Mai und Juni 2006 wurden Recherchen zu den Museumsauftritten durchgeführt. Ausgangspunkt war der Web-Museumsführer des Verbandes der Museen der Schweiz<sup>10</sup>. Dieser verzeichnete zu diesem Zeitpunkt 960 Museen.

Zu nahezu allen Museen finden sich Informationen im Netz. Über einen eigenständigen Internet-Auftritt verfügen 66 % der Einrichtungen während zu 32 % die Angaben zum Beispiel in den Seiten der Kommunen oder Tourismusbüros integriert sind.

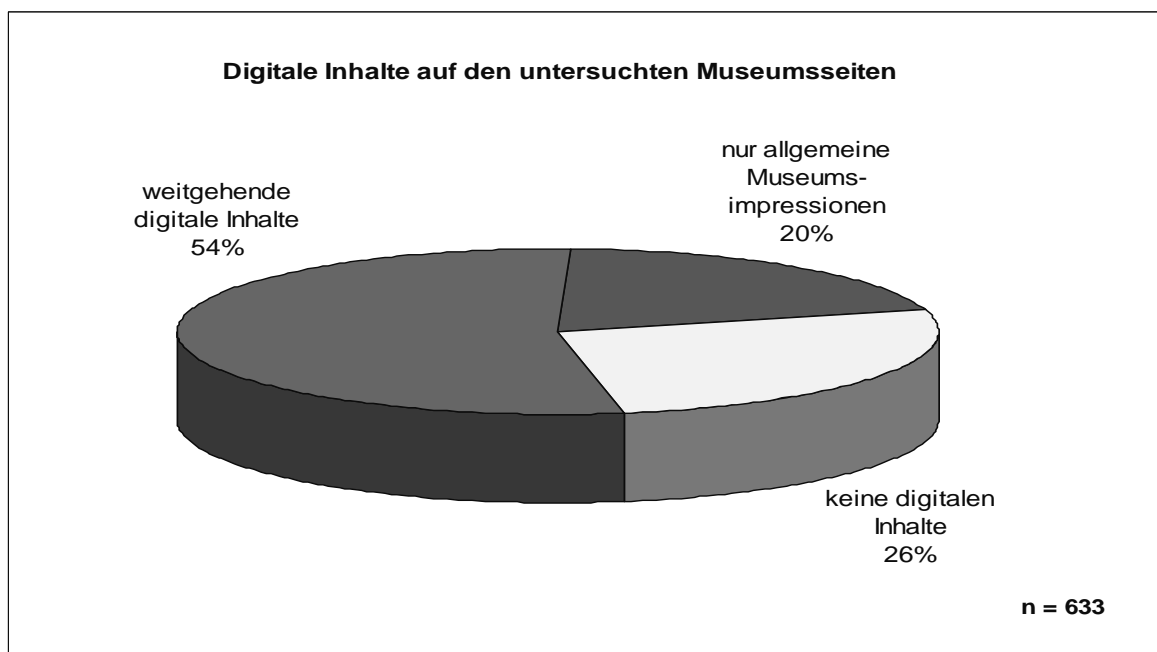
In einem nächsten Schritt wurden die selbstständigen Web-Auftritte (66 % bzw. 633 Museen) hinsichtlich ihres Umfangs untersucht. Rund die Hälfte offerieren umfangreiche Angebote. 20 % verfügen über ein erweitertes Angebot und rund 26 % sind nur mit den Basisinformationen vertreten. Diese Angaben beziehen sich meist auf Öffnungszeiten, Eintrittspreise, Beschreibungen der Anfahrtswege und eine mehr oder weniger ausführliche Beschreibung des Museums.

---

<sup>10</sup> Museumsführer des Verbandes der Museen der Schweiz <http://www.vms-ams.ch> Stand: 04.01.2007.



*Abbildung 7: Webpräsenz Schweizer Museen*



*Abbildung 8: Webpräsenz Schweizer Museen*

Die nachfolgende Übersicht verdeutlicht die allfälligen digitalen Angebotsformen. Am häufigsten ist der Museumsshop online präsent. Die Kommunikation mit der Presse erfolgt webgestützt. In vielfältiger Form finden sich Angebote, die dem virtuellen Besucher visuelle Eindrücke vermitteln. Hierzu zählen beispielsweise reali-

sierte Rundgänge als Text-/Bildfolge, Panoramen, Rundgang-Simulationen oder Pläne.

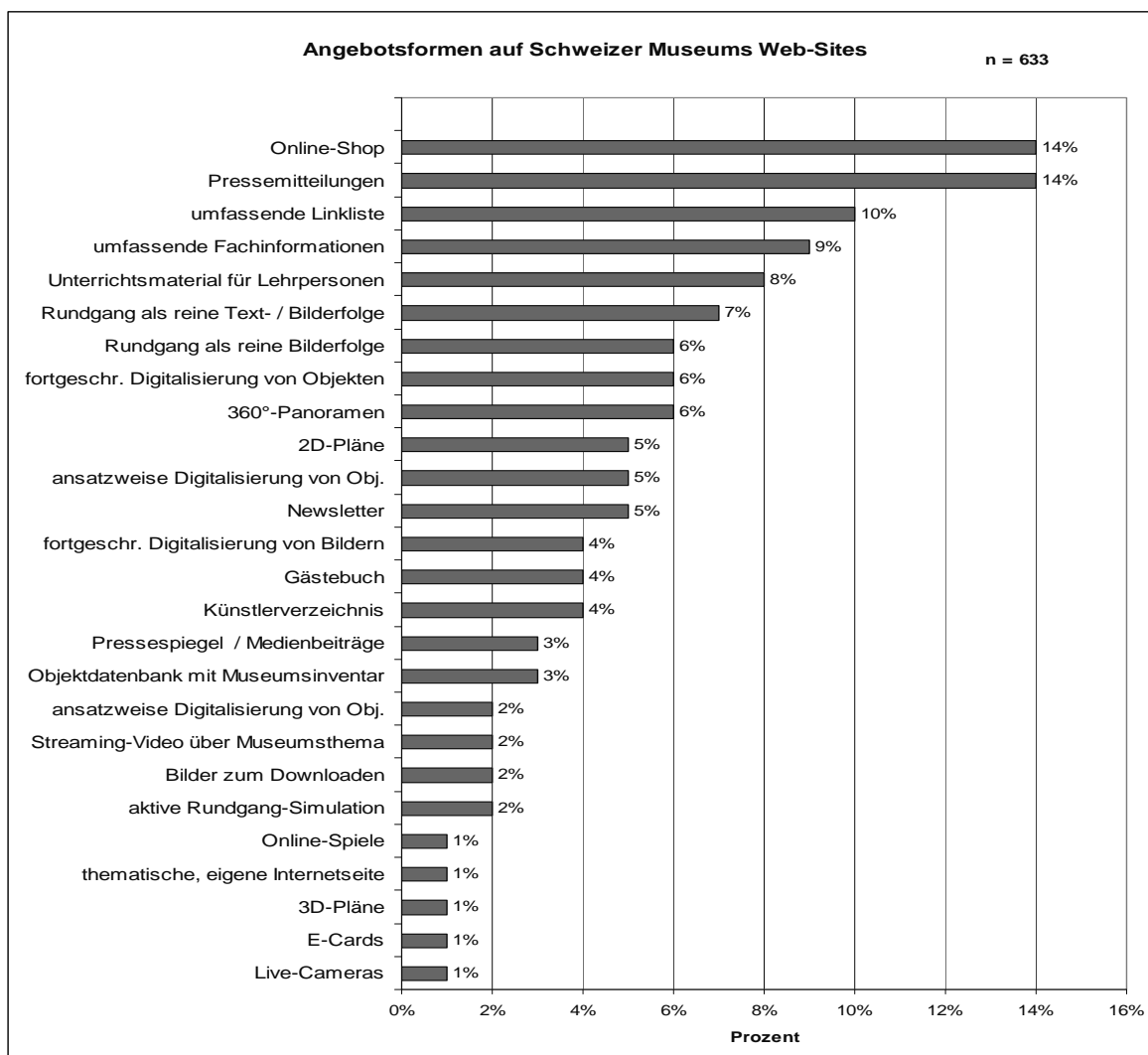


Abbildung 9: Angebotsformen

Viele Museen haben eine mehr oder minder große Anzahl von Ausstellungsobjekten digitalisiert. Mit einer ansatzweisen Digitalisierung werden hier Beispiele mit weniger als 20 Exponaten bezeichnet. Sind mehr als 20 Exponate in digitaler Form auf der Site erkennbar, wurde von einer fortgeschrittenen Digitalisierung ausgegangen. Insgesamt konnten 26 Künstlerverzeichnisse sowie 17 zugängliche Objektdatenbanken identifiziert werden.

### 3.3 Fazit

Erfreulich ist, dass die schweizerischen Museen nahezu vollständig im Netz vertreten sind. 66 % verfügen über einen eigenständigen Internetauftritt, während 33 % ihre Angaben auf den Seiten anderer Einrichtungen oder Dienste integriert haben.

Bei diesen handelt es sich meist um die kleineren Museen. Die Größe einer Einrichtung korreliert zumeist mit dem Grad des Einsatzes von Informationstechnologie.

Hinsichtlich der Digitalisierung der Objekte und des Virtualisierungsprozesses bestehen deutliche Defizite. Die Umfrage ergab, dass 71 % über digitale Daten verfügen und 56 % der Teilnehmer (140 Museen) ihre Objekte digital verzeichnen. Bei der Web-Recherche konnten jedoch nur 17 Objektdatenbanken und 26 Künstlerverzeichnisse nachgewiesen werden. Dies beschreibt deutlich die Diskrepanz zwischen intern existierenden und extern zugänglichen Informationsquellen. Der Bereich Standards in der Objekterfassung ist als kritisch einzuschätzen. Nur etwa 26 % der Museen verwenden etablierte Standards. Der Typus des Lernmuseums mit kontextorientierten Angeboten ist rar. Meta-Projekte mehrerer Einrichtungen sind kaum existent.

Diese Auszüge aus der Studie verdeutlichen, dass der Digitalisierungsgrad in den Museen der Schweiz noch nicht sehr weit vorgeschritten ist. Zurück liegen vor allem kleinere Museen mit bescheidener finanzieller Ausstattung. Für die Teilhabe an den skizzierten Entwicklungen im Kulturgüterbereich müssen die Rahmenbedingungen verbessert und die dargestellten Grundlagen auf breiterer Basis umgesetzt werden.

## 4 Quellen

- Bieber, Christina; Herget, Josef (2007). Stand der Digitalisierung im Museumsbereich in der Schweiz – Internationale Referenzprojekte und Handlungsempfehlungen. (Churer Schriften zur Informationswissenschaft, 14), Chur 2007 <http://www.iudchur.net/index.php?id=25> [04.01.2007].
- ICOM (1990). International Council of Museum. Code of Professional Ethics, Paris 1990.
- Hennings, Ralf-Dirk et al. (1996). Digitalisierte Bilder im Museum. Technische Tendenzen und organisatorisches Umfeld. Opladen 1996. (Berliner Schriften zur Museumskunde Band 14).
- Maier, Gerald (2005). Neue Wege zu digitalen Kulturgut-Informationen: Das BAM Portal: Gemeinsames Internetportal für Bibliotheken, Archive und Museum ([www.bam-portal.de](http://www.bam-portal.de).) In: *Bibliothek Forschung und Praxis*; Vol. 29, Nr. 2, 2005, 220-229.
- Niggemann, Elisabeth (2006). The European Digital Library – A Project of the Conference of European National Librarians. In: *ERICIM News (Special European Digital Library)* Vol. 66, Juli 2006, 15-16.
- O'Reilly, Tim (2005). What is Web 2.0. Design Pattern and Business Models for the Next Generation of Software. Version 09.03.2005. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> [04.01.2007].

- Rehm, Clemens (2006). Landesarchiv Baden-Württemberg diskutiert erste Ergebnisse zur Archivierung elektronischer Unterlagen. *Archivalia Blog*, 2006. Beitrag vom 13.10.2006. <http://archiv.twoday.net/stories/2809415/>.
- Schweibenz, Werner (1998). Das virtuelle Museum, Überlegungen zu Begriff und Erscheinungsformen des Museums im Internet. <http://www.compania-media.de/mum/mumfo>.
- Waidacher, Friedrich (1999). *Museologische Grundlagen der Objektdokumentation*. Berlin 1999.
- Wersig, Gernot; Schuck-Wersig, Petra: Deutsche Museen im Internet. Explorative Studie anlässlich des Workshops „Museumsbesuch im Multimedia-Zeitalter, wie werden die neuen Medien die Optionen der Museen verändern?“ vom 22. bis 23.05.1997 am Institut für Museumskunde, Berlin. <http://www.kommwiss.fu-berlin.de/forschung/vw/german/index.html>.
- BBS Program Catalogue <http://open.bbc.co.uk/catalogue/infax> [03.01.2006].
- European Digital Library, Projekthomepage: <http://www.edlproject.eu> [03.01.2007].
- Fine Arts Museum of San Francisco <http://www.thinker.org> [04.01.2007].
- Metropolitan Museum of Art <http://www.metmuseum.org> [04.01.2007].
- Museumsführer Schweiz <http://www.vms-ams.ch> [04.01.2007].
- PrestoSpace <http://www.prestospace.org/> [03.01.2007].
- SCRAN <http://www.scran.ac.uk> [04.01.2007].
- SWBplus, Datenbank des Südwestdeutschen Bibliotheksverbundes [http://www.bsz-bw.de/SWBplus/sp\\_recherche.shtml](http://www.bsz-bw.de/SWBplus/sp_recherche.shtml) [04.01.2007].
- Universitäts- und Landesbibliothek Düsseldorf, Online-Auskunft <http://www.ub.uni-duesseldorf.de/home/service/onlineauskunft> [04.01.2007].
- Video Active <http://videoactive.wordpress.com> [04.01.2007].



# Sceptical Doubts<sup>1</sup> Concerning Knowledge Capture within the Open Innovation Initiative\*

*Theodora Stathoulia*

Researcher in Information and Digital Technology

157 Aristotelous Street

GR-11251 Athens, Greece

*theodorastathoulia@yahoo.gr*

## Abstract

This paper presents an epistemic framework on knowledge management, theory, capture and representation within the *Open Innovation* domain. With view to this paper's objective that is to contribute to 'knowledge capture' the effort will be directed to the development of a different basis of dialogue. Thus, we argue that 'knowledge models' based on concept maps need supplementary considerations on knowledge theory. Our purpose to sign the closeness of today's understanding about knowledge and knowledge capture mechanisms to radical empiricism foundations represented by Hume's work is the basic epistemic standing point. Furthermore, the use of Hume's words on 'sceptical doubts' is due to our purpose to underline our observation that current bibliography is occupied with an intense empiricism. This work supports the idea that the system of knowledge that is for Leibniz a system of truths should be deductively based on the division and the analysis of concepts and symbolisms.

---

<sup>1</sup> Doubts is used here to underline epistemic grounds (in the way Stanford Encyclopedia of Philosophy defines doubts in philosophy, <http://plato.stanford.edu/>)

[The title of the paper purposefully uses the title of section IV: 'Sceptical Doubts Concerning the Operations of Understanding' of Hume's book 'Enquiries are concerning the Human understanding', reprinted from the posthumous edition of 1977, and edited with introduction, comparative tables of contents, and analytical index by L.A. Selby-Bigge. (The Project Gutenberg eBook of 'A Enquiry Concerning Human Understanding', by D. Hume). This is due to our purpose to sign the closeness of today's understanding about knowledge and knowledge capture mechanisms to radical empiricism foundations represented by Hume's work].

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 237-241.

## I Introduction

It is in our common tradition to believe that the production and the disposal of knowledge hold a supreme role in the advances of science, research, and business. With respect to the conference's framework questioning whether *Open Innovation* is defined within New Perspectives in the Context of Information and Knowledge, our sceptical doubts come to confirm the manifold nature of the theme.

Considering ideas which proceed theoretical imperatives such as “the basic economic resource is no longer capital, or natural resources, or labour, but is, and will be knowledge. This basic tenet defines today's enterprise richest resource to be the knowledge and ideas residing in the minds of a particular organization's employees, customers, and vendors” [Becerra-Fernandez 98], the understanding and the advancement of the field seems to be of great value. The citation mentioned above is exemplified by practices underlying institutional impact on the exploitation of knowledge mechanisms such as those accomplished by the *Center for Innovation and knowledge management* which assists “companies that are in business of developing knowledge management tools, such as groupware, collaborative computing, document management, [...]”.[Becerra-Fernandez 98].

Consequently, the consolidation concerning theoretical enquiries to knowledge exceeds the discussion to technological reserves. We are by no means emphatic to theories and technological reserves inconsistencies. We do not either aim to present aspects of knowledge management including the technological practicalities.

## 2 Discussion

Our enquiry focuses on knowledge theory, knowledge capture and representation. Does this imply the necessity to shift to a new paradigm in the way the scientific legacy has introduced ‘paradigm change’ to science? It is precipitate to be stated and also, beyond the scope of this work. With view to this paper's objective, that is to contribute to ‘knowledge capture’ within the *Open Innovation* domain, the effort will be directed to the development of a different basis of dialogue. Thus, we should guide our thoughts to the establishment of ideas, concepts and analysis.

For instance, ‘Knowledge models’ [Hayes 05] have been successfully mapped and have captured knowledge within institutional lives. But should we deny that ‘knowledge models’ based on concept maps need supplementary considerations on

knowledge theory? Some of them has been attempted to be framed in the following passage.

What about the incapacities of languages ‘especially the ordinary languages of common life due to its preoccupation with the sense world and its consequent vagueness on ultimate matters’ [Morrow 29]. Again, are ‘concept maps’ the ‘ultimate matters’? Here, then, ‘Knowledge models’ should be examined within the doctrine of philosophy to express the foundations of language and ontology.

Are Kaminsky’s ‘ontological commitments’ [Frye 71] enough to represent the ‘ultimate matters’? Furthermore, the ontological commitment to subject as predicate sentence structure and fundamental in languages is part of knowledge capture and representation, or business technical language does not need such commitments?

Someone might argue that empirical concepts easily extracted from knowledge business communication channels are sufficient in describing the elementary in business resources.

From the very beginning of this presentation, the use of Hume’s words on ‘sceptical doubts’ is due to our purpose to underline our observation that current bibliography is occupied with an intense empiricism that resides in the idea that knowledge lies among or/and within firms and it consolidates a practice of research concerning integrated knowledge in networks, which in sequence have been developed by knowledge capture.

Alternatively, one might draw attention to considerations on knowledge capture in more advanced environments that are getting a solid ground of references to ‘communities of practice’ and to ‘knowledge networks’ [Merali and Davies 01], and provide the certainty of knowledge theories prosperity. The use of the term ‘knowledge capture’ indicates that we have already placed in orbit the solar bodies round the observer (in opposition to the evolutionary theory of Copernicus who wanted the observer to move round the solar bodies) and obscures our view on the tradition of rationalism, on Kantian *a priori* Knowledge, on Leibnizian objective idealism, that time and space are ‘orders’ and ‘relations’, not entities or existences [Cassirer 43]. How are we going to capture and represent ‘relations’? *Semantic schemata* are the answer? However, they are not enough unless knowledge is identified within its conceptual context deriving from the historical background, the linguistic parameters, and why not, its controversies as well as its own *a priori* existence.

With respect to this principle, ontology, either Newtonian or Leibnizian [Brewster and O’Hara 04], with the former to refer to views on ontology as finding elegant

simplifying principles and the later to refer to the extreme complexity of experience, knowledge representation, has not succeeded in interpreting the actual world.

Any of these approaches might be questioned, e. g. which is the base of their arguments? In other words, where do we stand in order to build our question over knowledge capture? Our argument traces its roots back in knowledge theories. It is by no means a research proposal. It is an epistemic issue. The vehicle to this journey to knowledge theory is the tradition of the two schools, the empiricist and the rationalist. Much of Hume and Hume's belief that we expect similar effects from apparently similar causes permits us to assume that knowledge is the ultimate formula of a matter of fact, a reasoning based entirely on experience, as it is recognized in contemporary linear knowledge models. This is to say, a greater amount of cases from the actual business world will produce the effects we need for an effective response to socioeconomic changes. But we already know that so far this response hasn't been proved adequate in representing the scientific work. As far as rationalist thought is concerned, we know that Kant wrote his *Critique* mainly against Hume and Leibniz. On the other hand, rationalist thought, with Kant to write his *Critique* first in opposition to Hume and to Leibniz, for different reasons, although he has been accused [Scruton 81] as a Leibnizian or a 'Prussian Hume', in his theory of understanding turns the attention not to the origins of experience but to what lies in it. It is the analysis of concepts and the limits of understanding which build the line of his thought.

### 3 Conclusions

In order to line around the boundaries of our argument, a rough outline of the two schools of thoughts on knowledge theory was attempted above. Thus, the foundation of our proposal will be the reconciliation of the empiric supervisory knowledge with the purely conceptual. The system of knowledge that is for Leibniz a system of truths should be deductively based on the division and the analysis of concepts. The ideal of knowledge representation should be supported by the Kantian deductive method. Thus our knowledge schemata will derive from a system of truths universally accepted in the classic hierarchical representation of the scientific knowledge. We have a long way ahead to connect, however not ontologically, the two schools of thoughts, the empiricist and the rationalist, and moreover to find the structures and the functions of languages we need to serve our knowledge representation.

The reconciliation of the ‘observer’ with the ‘object’, the ‘harmony’ between the capacities of the knower and the nature of the known, the Kantian ‘transcendental idealism’, after all the existence following the *a priori* knowledge serves fundamental aspects of our proposition in this work.

Conclusively, to this extent we can answer affirmatively to the initial question if there is a new perspective in the context of information and knowledge within the *Open Innovation*. However, the new perspective should be examined by the replacement of facts by more “symbolism in thoughts” [Cassirer 43]. Now, we need more concepts and syllogism and less pragmatism. This prospect is supposed to bring in the surface the topics all along with the forms that will be addressed in a new landscape that is shaped by the needs emerging in Knowledge management, especially those regarding the unification of concepts and case-based experiences (external knowledge), and corporate fact-based knowledge (implicit knowledge).

## References

- [Becerra-Fernandez 98] I. Becerra-Fernandez, ‘Center for Innovation and Knowledge Management’, *SIGGROUP Bulletin*, vol. 19, no. 1, Apr. 1998, pp.46-52.
- [Hayes 05] Hayes, P. et. al., ‘Collaborative Knowledge capture in Ontologies’, *K-CAP’05*, October 2 – 5, 2005, p.99-106 [Ford et al. [11] used the term *knowledge models* to denote groups of interwoven concept maps and associated resources. Knowledge models have been developed using CmapTools [3] for several large institutional memory and expert knowledge preservation tasks, including launch vehicle systems integration [5], mesoscale weather forecasting [13], Thai fabric design, Mars exploration [2], and nuclear power air effluent analysis [6]. Knowledge acquisition using concept maps is also very efficient, with production averaging two useful propositions generated per session minute].
- [Morrow 29] G. Morrow, ‘The theory of Plato’s seventh epistle’, *The Philosophical Review*, vol. 38 (4), 1929, pp. 326-349.
- [Frye 71] M. Frye (Book Review) ‘Language and Ontology’ by J. Kaminsky, *The Philosophical Review*, vol. 80 (3), July 1971, pp.394-396.
- [Merali and Davies 01] Y. Merali and J. Davies, ‘Knowledge Capture and Utilization in Virtual Communities’, *K-CAP’01*, October 22-23, 2001, pp 92-99.
- [Cassirer 43] E. Cassirer, ‘Newton and Leibniz’, *The Philosophical Review*, vol. 52(4), July 1943, pp. 366-391.
- [Brewster and O’Hara 04] C. Brewster and K. O’ Hara, ‘Knowledge representation with ontologies: the present and the Future’, *IEEE, Intelligent Systems*, January/February 2004, pp. 72-81
- [Scruton 81] R. Scruton, ‘*Kant*’, Oxford University Press, 1981.



# Tagging, Folksonomy & Co – Renaissance of Manual Indexing?\*

*Jakob Voß*

Common Library Network GBV  
Göttingen, Germany  
*[jakob.voss@gbv.de](mailto:jakob.voss@gbv.de)*

## **Abstract**

This paper gives an overview of current trends in manual indexing on the Web. Along with a general rise of user generated content there are more and more tagging systems that allow users to annotate digital resources with tags (keywords) and share their annotations with other users. Tagging is frequently seen in contrast to traditional knowledge organization systems or as something completely new. This paper shows that tagging should better be seen as a popular form of manual indexing on the Web. Difference between controlled and free indexing blurs with sufficient feedback mechanisms. A revised typology of tagging systems is presented that includes different user roles and knowledge organization systems with hierarchical relationships and vocabulary control. A detailed bibliography of current research in collaborative tagging is included.

## **1 Introduction**

The World Wide Web, a framework originally designed for information management [Berners-Lee 1989], has long ago become a heterogeneous, exponentially growing mass of connected, digital resources. After first, unsuccessful attempts to classify the Web with traditional, intellectual methods of library and information science, the standard to search the Web is now fulltext indexing – most of all made popular by Google's PageRank algorithm. The success of such automatic techniques is a reason why "many now working in information retrieval seem completely unaware that procedures other than fully automatic ones have been applied, with some

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 243-254.

success, to information retrieval for more than 100 years, and that there exist an information retrieval literature beyond that of the computer science community.” [Lancaster 2003] However in the recent years there is a renaissance of manual subject indexing and analysis: Structured metadata is published with techniques like RSS, OAI-PMH, and RDF. OpenSearch<sup>1</sup> and browser search plug-ins allow it to aggregate specialised search engines. Last but not least many popular social software systems contain methods to annotate resources with keywords. This type of manual indexing is called tagging with index terms referred to as tags. Based on [Marlow et al. 2006] this paper presents a revised typology of tagging systems that also includes systems with controlled and structured vocabularies. Section 2 gives a short introduction to current tagging systems and its research. Afterwards (section 3) theory of subject indexing is pictured with the indexing process, typology of knowledge organization systems, and an unconventional look at vocabulary control. In section 4 the typology of tagging systems is presented with conclusion in section 5.

## 2 Tagging systems on the rise

Tagging is referred to with several names: collaborative tagging, social classification, social indexing, folksonomy etc. The basic principle is that end users do subject indexing instead of experts only, and the assigned tags are being shown immediately on the Web. The number of websites that support tagging has rapidly increased since 2004. Popular examples are del.icio.us (<http://del.icio.us>), furl (<http://furl.net>), reddit (<http://reddit.com>), and Digg (<http://digg.net>) for bookmarks [Hammond et al. 2005] and flickr (<http://flickr.com>, [Winget 2006]) for photos. Weblog authors usually tag their articles and specialized search engines like Technorati (<http://technorati.com/>) and RawSugar (<http://rawsugar.com>) make use of it. But tagging is not limited to simple keywords only: BibSonomy (<http://bibsonomy.org>, [Hotho et al. 2006b,a]), Connotea (<http://connotea.org>, [Lund et al. 2005]), CiteU-Like (<http://citeulike.org/>), and LibraryThing (<http://librarything.com>) allow users to manage and share bibliographic metadata on the Web (also known as social reference managing or collaborative cataloguing). In addition to librarian’s subject indexing the University of Pennsylvania Library allows users to tag records in their online catalog since 2005 (<http://tags.library.upenn.edu/>). Other systems to tag bibliographic data are LibraryThing (<http://www.librarything.com>) and Amazon’s tagging feature (<http://amazon.com/gp/tagging/cloud/>). The popular free encyclopaedia

---

<sup>1</sup> To gain an insight on RSS, OAI-PMH, RDF, and OpenSearch see <http://en.wikipedia.org/>.



Wikipedia contains so called categories that are used as hierarchical tags to order the articles by topic [Voss 2006]. Apart from social software there is also a rise of manual indexing in other fields [Wright 2005, Maislin 2005].

The details of tagging vary a lot but all applications are designed to be used as easy and as open as possible. Sometimes the greenness in theory of users and developers let you stumble upon known problems like homonyms and synonyms but on the other hand unloaded trial and error has led to many unconventional and innovative solutions.

## **2.1 Research on Tagging**

The astonishing popularity of tagging led some even claim that it would overcome classification systems [Shirky 2005], but it is more likely that there is no dichotomy [Crawford 2006] and tagging just stresses certain aspects of subject indexing. Meanwhile serious research about collaborative tagging is growing — hopefully it will not have to redo all the works that has been done in the 20<sup>th</sup> century. At the 15<sup>th</sup> World Wide Web Conference there was a Collaborative Web Tagging Workshop<sup>2</sup>. The 17<sup>th</sup> SIG/CR Classification Research Workshop was about Social Classification<sup>3</sup>. One of the first papers on folksonomies is [Mathes 2004]. Shirky's paper [Shirky 2005] has reached huge impact. It is probably outdated but still worth to read. A good overview until the beginning of 2006 is given in [Macgregor & McCulloch 2006]. Some papers that deal with specific tagging systems are cited at the beginning of section 2. Trant and Smith describe the application of tagging in a museum [Trant & Wyman 2006, Trant 2006, Smith 2006]. Other works focus on tagging in enterprises [Farrell & Lau 2006, John & Seligmann 2006, Laurie Damianos 2006, Millen et al. 2006] or knowledge management [Wu et al. 2006]. Another application is tagging people to find experts [Toine Bogers & van den Bosch 2006, Farrell & Lau 2006]. Mathematical models of tagging are elaborated in [Tošić & Milićević 2006, Lambiotte & Ausloos 2005]. The usual model of tagging is a tripartite graph, the nodes being resources, users, and tags [Lambiotte & Ausloos 2005]. Several papers provide statistical analysis of tagging over time and evolution of tagging systems [Kowatsch et al. 2007, Cattuto et al. 2006, Dubinko et al. 2006, Hotho et al. 2006d, Lin et al. 2006]. Tagging behaviour is also topic of Kipp and [Campbell Kipp & Campbell 2006] and Feinberg [Feinberg 2006]. Types of structured and compound tagging are analyzed in [Bar-Ilan et al. 2006, Tonkin 2006, Guy & Tonkin 2006]. Like in traditional scientometrics you can find com-

---

<sup>2</sup> <http://www.rawsugar.com/www2006/taggingworkshopschedule.html>

<sup>3</sup> <http://www.slais.ubc.ca/users/sigcr/events.html>

munities and trends based on tagging data [Jäschke et al. 2006, Hotho et al. 2006d]. Voss [Voss 2006] finds typical distributions among different types of tagging systems and compares tagging systems with traditional classification and thesaurus structures. Tennis [Tennis 2006] uses framework analysis to compare social tagging and subject cataloguing. Tagging is manual indexing instead of automatic indexing. Ironically a focus of research is again on automatic systems that do data mining in tagging data [Aurnhammer et al. 2006, Hotho et al. 2006b,c, Schmitz 2006, Schmitz et al. 2006]. Heymann and Garcia-Molina [Heymann & Garcia-Molina 2006] presented an algorithm to automatically generate hierarchies of tags out of flat, aggregated tagging systems with del.icio.us data. Similar approaches are used by Begelman et al. [Begelman et al. 2006] and Mika et al. [Mika 2005]. Research on tagging mostly comes from computer science and library science — obviously there is a lack of input from psychology, sociology, and cognitive science in general (an exception from philosophy is Campell [Campbell 2006] who applies Husserl’s theory of phenomenology to tagging).

### 3 The indexing process

The main purpose of subject indexing is to construct a representation of a resource that is being tagged. According to Lancaster [Lancaster, 2003, chapter 2], subject indexing involves two steps: conceptual analysis and translation (see figure 1).

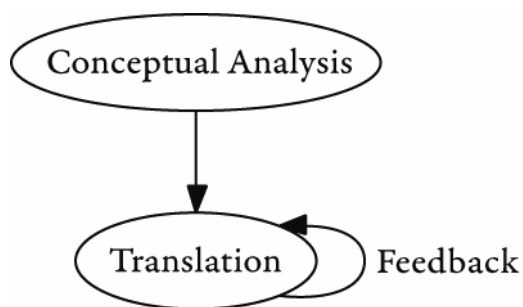


Figure 1: Steps in subject indexing

These are intellectually separate although they are not always clearly distinguished. The semiotic triangle can be applied to indexing to demonstrate the distinction between object (resource), concept (what the resource is about), and symbol (set of tags to represent the resource). Conceptual analysis involves deciding on what a resource is about and what is relevant in particular. Note that the result of conceptual analysis heavily depends on the needs and interests of users that a resource is tagged for — different people can be interested in different aspects. Translation is the process of finding an appropriate set of index terms (tags) that represent the substance of

conceptual analysis. Tags can be extracted from the resource or assigned by an indexer. Many studies have shown that high consistency among different indexers is very difficult to achieve and affected by many factors [Lancaster, 2003, chapter 3]. One factor is control of the vocabulary that is used for tagging. Synonyms (multiple words and spellings for the same concept) and homonyms/homographs (words with different meanings) are frequent problems in the process of translation. A controlled vocabulary tries to eliminate them by providing a list of preferred and non-preferred terms, often together with definitions and a semantic structure. Controlled vocabularies are subsumed as knowledge organization systems (KOS) [Zeng & Chan 2004]. These systems have been studied and developed in library and information science for more than 100 years. Popular examples are the Dewey Decimal Classification, Ranganathan's faceted classification, and the first thesauri in the 1960s. Beginning with the 1950s library and information science has lost its leading role in the development of information retrieval systems and a rich variety of KOS has come into existence. However it is one of the constant activities of this profession to summarize and evaluate the complexity of attempts to organize the world's knowledge.

### **3.1 Typology of knowledge organization systems**

Hodge, Zeng, and Tudhope [Hodge 2000, Zeng 2000, Tudhope 2006] distinguish by growing degree of language control and growing strength of semantic structure: term list, classifications and categories, and relationship groups. Term lists like authority files, glossaries, gazetteers, and dictionaries emphasize lists of terms often with definitions. Classifications and categories like subject headings and classification schemes (also known as taxonomies) emphasize the creation of subject sets. Relationships groups like thesauri, semantic networks, and ontologies emphasize the connections between concepts. Apart from the training of what now may be called ontology engineers the theoretical research on knowledge organization systems has had little impact on technical development. Only now common formats are being standardized with SKOS<sup>4</sup>, the microformats movement<sup>5</sup> and other initiatives. Common formats are a necessary but not sufficient condition for interoperability among knowledge organization systems — an important but also frequently underestimated task [Zeng & Chan 2004, Mayr 2006].

---

<sup>4</sup> <http://www.w3.org/2004/02/skos/>

<sup>5</sup> <http://microformats.org/>

### 3.2 Vocabulary control and feedback

In the process of indexing the controlled vocabulary is used to supply translation via feedback (figure 1). The indexer searches for index terms supported by the structure of the knowledge organization systems until he finds the best matching tag. Also search is supported by the structure of the knowledge organization systems. Collaborative tagging also provides feedback.

A special kind of tagging system is the category system of Wikipedia. The free encyclopaedia is probably the first application of collaborative tagging with a thesaurus [Voss 2006]. The extend of contribution in Wikipedia is distributed very inhomogeneously (more precise it is a power law [Voss 2005]) – this also applies for the category system. Everyone is allowed to change and add categories but most authors only edit the article text instead of tagging articles and even less authors change and add the category system. Furthermore each article is not tagged independently by every user but users have to agree on a single set of categories per article. So tagging in Wikipedia is somewhere between indexing with a controlled vocabulary and free keywords. Most of the time authors just use the categories that exist but they can also switch to editing the vocabulary at any time. The emerging system may look partly chaotic but rather useful. With a comparison of Wikipedia and the AGRO-VOC<sup>6</sup> thesaurus Milne et al. [Milne et al. 2006] show that domain-specific thesauri can be enriched and created with Wikipedia's category and link structure.

We can deduce that the border between free keyword tagging on the one hand and tagging with a controlled vocabulary is less strict. Although most tagging systems do not implement vocabulary control there is almost always some feedback that influences tagging behaviour towards consensus: the Folksonomy emerges [Mathes 2004]. This phenomena is also known as emergent semantics or Wisdom of the crowds [Surowiecki 2004] (But you should keep in mind that masses do not always act wise – see Lanier's critic of 'Digital Maoism' [Lanier 2006]).

## 4 Typology of tagging systems

Based on Marlow's taxonomy of tagging systems [Marlow et al. 2006] I provide a revised typology. The following key dimensions do not represent simple continuums but basic properties that should be clarified for a given tagging system — so they are presented here as questions. The best corresponding dimension of Marlow's typology is noted in brackets.

---

<sup>6</sup> <http://www.fao.org/agrovoc>

**Tagging Rights** (Tagging Rights)

Who is allowed to tag resources? Can any user tag any resource or are there restrictions? Are restrictions based on resources, tags, or users? Who decides on restrictions? Is there a distinction between tags by different types of users and resources?

**Source of Resource** (Source of material)

Do users contribute resources and have resources been created or just supplied by users? Or do users tag resources that are already in the system? Who decides which resources are tagged?

**Representation of Resource** (Type of object)

What kind of resource is being tagged? How are resources presented while tagging (for instance autopsy principle)?

**Tagging Feedback** (Tagging Support)

How does the interface support tag entry? Do users see other tags assigned to the resource by other users or other resources tagged with the same tags? Does the system suggest tags and if so based on which algorithms? Does the system reject inappropriate tags?

**Tag Aggregation** (Aggregation)

Can a tag be assigned only once to a resource (set model) or can the same tag can be assigned multiple times (bag-model with aggregation)?

**Vocabulary control** ( – )

Is there a restriction on which tags to use and which tags not to use? Are tags created while tagging or is management of the vocabulary a separated task? Who manages the vocabulary, how frequently is it updated, and how are changes recorded?

**Vocabulary Connectivity** ( – )

Are tags connected with relations? Are relations associative (authority file), monohierarchical (classification or taxonomy), multihierarchical (thesaurus), or typed (ontology)? Where do the relations come from? Are relations limited to the common vocabulary (precoordination) or can they dynamically be used in tagging (postcoordination with syntactic indexing)?

**Resource Connectivity** (Resource Connectivity)

How are resources connected to each other with links or grouped hierarchically? Can resources be tagged on different hierarchy levels? How are connections created?

### **Automatic Tagging ( – )**

Is tagging enriched with automatically created tags and relations (for instance file types, automatic expansion of terms etc.)?

Marlow also includes social connectivity, but only the influence of this property on other dimensions like Tagging Feedback and Source of Resources is of interest. The analysis shows that the classic tripartite model of tagging with resources, users, and tags is too simplified to cover the variety of tagging system. Depending on the application you can distinguish different kind of resources, tags, and users. At least you should distinguish four user roles:

**Resource Author:** A person that creates or edits a resource

**Resource Collector:** A person that adds a resource to a tagging system

**Indexer or Tagger:** A person that tags resources

**Searcher:** A person that uses tags to search for resources

In a system that uses controlled vocabularies, the role of a Vocabulary Editor (a person that maintains the vocabulary) may be added. In most systems some of the roles overlap and people can fulfil different roles at different times (a large library with strict distribution of tasks may be a counterexample). For instance the author of a private blog combines 1, 2, and 3, a user of del.icio.us combines either 2 and 3 (tagging a new webpage) or 3 and 4 (copying a webpage that someone else has already tagged), and a Wikipedia author combines either 1 and 2 (new articles) or 1 and 3 (existing articles). Existing applications of tagging can be described and classified with the typology above and the four roles. For instance author keywords are assigned by the author of a publication (Tagging Rights, Source of Resources, Representation of Resource), with no or little feedback (Tagging Feedback), other people may add their own keywords, but there is little or no aggregation (Tag Aggregation), keywords are free and unconnected (Vocabulary Control, Vocabulary Connectivity).

## **5 Conclusion**

The popularity of collaborative tagging on the Web has resurged interest in manual indexing. Tagging systems encourage users to manually annotate digital objects with free keywords and share their annotations. Tags are directly assigned by anyone who likes to participate. The instant visibility is motivation and helps to install feedback mechanism. Through feedback the drawbacks of uncontrolled indexing are less dramatic than in previous systems and the border between controlled and free in-

dexing starts to blur. Vocabulary control and relationships between index terms should not be distinctive features of tagging systems and traditional knowledge organization systems but possible properties of manual indexing systems. Further research is needed to find out under which circumstances which features (for instance vocabulary control) are needed and how they influence tagging behaviour, evolution of the tagging system, and benefits of tagging for searching and browsing. The typology of tagging systems and roles that was presented in section 4 combines all of them. The possibility to allow non-experts to add keywords has made collaborative tagging so popular — but it is nothing fundamentally new. Perhaps the most important feature of tagging systems on the Web is its implementation or how Joseph Busch entitled his keynote speech at the ASIST SIG-CR workshop: “It’s the interface, stupid!” Today’s tagging websites make many traditional knowledge organization systems look like Stone Age technique: effective but just too uncomfortable. Some of the costly created thesauri and classifications are not even accessible in digital form at all (because of licensing issues grounded in a pre-digital understanding of copyright or because of a lack of technological skills)! But also computer scientists tend to forget that a clever interface to support tagging can be worth much more than any elaborated algorithm. Anyway the art of creating interfaces for developed tagging systems is still in its infancy. Knowledge organization will always need manual input so it is costly to manage — but Wikipedia showed that groups of volunteers can create large knowledge resources if a common goal and the right toolkit exist! And obviously there is not one way of indexing that fits for all applications. Collaborative Tagging is neither the successor of traditional indexing nor a short-dated trend but — like Tennis [Tennis 2006] concludes — a catalyst for improvement and innovation in indexing.

## 6 References

- Aurnhammer, M., Steels, L., & Hanappe, P. (2006). Integrating collaborative tagging and emergent semantics for image retrieval. In *Proceedings of the WWW 2006 Collaborative Web Tagging Workshop*.
- Bar-Ilan, J., Shoham, S., Idan, A., Miller, Y., & Shachak, A. (2006). Structured vs. unstructured tagging: A case study. In *Proceedings of the WWW 2006 Collaborative Web Tagging Workshop*.
- Begelman, G., Keller, P., & Smadja, F. (2006). Automated tag clustering: Improving search and exploration in the tag space. In *Proceedings of the WWW 2006 Collaborative Web Tagging Workshop*.
- Berners-Lee, T. (1989). *Information Management: A Proposal*. Technical report, CERN.
- Campbell, D. G. (2006). A phenomenological framework for the relationship between

- the semantic web and user-centered tagging systems. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.
- Cattuto, C., Loreto, V., & Pietronero, L. (2006). Collaborative tagging and semiotic dynamics.
- Crawford, W. (2006). Folksonomy and Dichotomy. *Cites & Insights*, 6 (4).
- Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., & Tomkins, A. (2006). Visualizing tags over time. In Proceedings of the 15<sup>th</sup> Conference on World Wide Web.
- Farrell, S. & Lau, T. (2006). Fringe contacts: People-tagging for the enterprise. In Proceedings of the WWW 2006 Collaborative Web Tagging Workshop.
- Feinberg, M. (2006). An examination of authority in social classification systems. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.
- Guy, M. & Tonkin, E. (2006). Folksonomies tidying up tags? *D-Lib Magazine*, 12.
- Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social Bookmarking Tools (I): A General Review. *D-Lib Magazine*, 11.
- Heymann, P. & Garcia-Molina, H. (2006). Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical report, InfoLab, Stanford.
- Hodge, G. (2000). Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. CLIR Publications.
- Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006a). Das entstehen von semantik in bibsonomy. In *Social Software in der Wertschöpfung*. Nomos.
- Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006b). Emergent semantics in bibsonomy. In Proceedings of the 1st Workshop on Applications of Semantic Technologies.
- Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006c). Information Retrieval in Folksonomies: Search and Ranking. In Proceedings of the 3rd European Semantic Web Conference, volume 4011 of Lecture Notes in Computer Science.
- Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006d). Trend Detection in Folksonomies. In Proceedings of the 1st Conference on Semantics And Digital Media Technology.
- Jäschke, R., Hotho, A., Schmitz, C., & Stumme, G. (2006). Wege zur Entdeckung von Communities in Folksonomies. In Proceedings des 18. Workshop Grundlagen von Datenbanken, (pp. 80 – 84).
- John, A. & Seligmann, D. (2006). Collaborative tagging and expertise in the enterprise. In Proceedings of the WWW 2006 Collaborative Web Tagging Workshop.
- Kipp, M. & Campbell, G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In Proceedings of the ASIST Annual Meeting.
- Kowatsch, T., Muenster, T., & Maass, W. (2007). Social Economics and the Shift of Common Interests.
- Lambiotte, R. & Ausloos, M. (2005). Collaborative tagging as a tripartite network. *Lecture Notes in Computer Science*, 3993, 1114 – 1117.
- Lancaster, F. W. (2003). *Indexing and Abstracting in Theory and Practise* (Third edition ed.). facet publishing.



- Lanier, J. (2006). Digital maoism: The hazards of the new online collectivism. Edge.
- Laurie Damianos, John Griffith, D. C. (2006). Onomi: Social bookmarking on a corporate intranet. In Proceedings of the WWW 2006 Collaborative Web Tagging Workshop.
- Lin, X., Beaudoin, J. E., Bui, Y., & Desai, K. (2006). Exploring characteristics of social classification. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.
- Lund, B., Hammond, T., Flack, M., & Hannay, T. (2005). Social Bookmarking Tools (II): A Case Study -Connotea. D-Lib Magazine, 11.
- Macgregor, G. & McCulloch, E. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. Library Review, 55, 291 – 300.
- Maislin, S. (2005). The indexing revival. intercom, 6 – 8.
- Marlow, C., Naaman, M., Boyd, D., & Davis, M. (2006). Position paper, tagging, taxonomy, flickr, article, toread. In Proceedings of the WWW 2006 Collaborative Web Tagging Workshop.
- Mathes, A. (2004). Folksonomies -Cooperative Classification and Communication Through Shared Metadata. Technical report, Graduate School of Library and Information Science, Illinois Urbana-Champaign.
- Mayr, P. (2006). Thesauri, Klassifikationen & Co – die Renaissance der kontrollierten Vokabulare? In Vom Wandel der Wissensorganisation im Informationszeitalter (pp. 151 – 170). Bock und Herchen.
- Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics. In Proceedings of the 4<sup>th</sup> International Semantic Web Conference.
- Millen, D. R., Feinberg, J., & Kerr, B. (2006). Social bookmarking in the enterprise. In Proceedings of the SIGCHI conference on Human Factors in computing systems.
- Milne, D., Medelyan, O., & Witten, I. H. (2006). Mining Domain-Specific Thesauri from Wikipedia: A case study. In Proceedings of the Conference on Web Intelligence.
- Schmitz, C., Hotho, A., Jäschke, R., & Stumme, G. (2006). Mining Association Rules in Folksonomies. In Proceedings of the 10<sup>th</sup> IFCS Conference.
- Schmitz, P. (2006). Inducing ontology from flickr tags. In Proceedings of the WWW 2006 Collaborative Web Tagging Workshop.
- Shirky, C. (2005). Ontology is overrated.
- Smith, M. K. (2006). Viewer tagging in art museums: Comparisons to concepts and vocabularies of art museum visitors. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.
- Surowiecki, J. (2004). The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations. Little Brown.
- Tennis, J. T. (2006). Social tagging and the next steps for indexing. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.
- Toine Bogers, W. T. & van den Bosch, A. (2006). Expertise classification: Collaborative classification vs. automatic extraction. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.

- Tonkin, E. (2006). Searching the long tail: Hidden structure in social tagging. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.
- Tošić, M. & Milićević, V. (2006). The Semantics of Collaborative Tagging System. In Proceedings of the 2nd Workshop on Scripting for the Semantic Web.
- Trant, J. (2006). Social classification and folksonomy in art museums: Early data from the steve.museum tagger prototype. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.
- Trant, J. & Wyman, B. (2006). Investigating social tagging and folksonomy in art museums with steve.museum. In Proceedings of the WWW 2006 Collaborative Web Tagging Workshop.
- Tudhope, D. (2006). A tentative typology of KOS: towards a KOS of KOS? In Proceedings of the 5<sup>th</sup> European NKOS Workshop.
- Voss, J. (2005). Measuring wikipedia. In Proceedings of the 10<sup>th</sup> Conference of the International Society for Scientometrics and Informetrics.
- Voss, J. (2006). Collaborative thesaurus tagging the wikipedia way. Technical report, Wikimetrics.
- Winget, M. (2006). User-defined classification on the online photo sharing site flickr. In Proceedings of the 17<sup>th</sup> SIG Classification Research Workshop.
- Wright, I. (2005). The future of indexing? WritersUA.
- Wu, H., Zubair, M., & Maly, K. (2006). Harvesting social knowledge from folksonomies. In Proceedings of the 17<sup>th</sup> conference on Hypertext and hypermedia.
- Zeng, M. L. (2000). Taxonomy of knowledge organization sources/systems.
- Zeng, M. L. & Chan, L. M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55 (5), 377 – 395.

# Feedback zur Beförderung netzbasierter Wissenskommunikation in K<sub>3</sub>\*

*Joachim Griesbaum, Wolfgang Semar, Tao Jiang &  
Rainer Kuhlen*

Universität Konstanz  
Informationswissenschaft  
78457 Konstanz  
Deutschland

*joachim.griesbaum@inf.uni-konstanz.de*

## **Zusammenfassung**

In diesem Beitrag werden, mit Schwerpunkt auf Feedback, Verfahren zur Beförderung netzbasierter Kooperation in Lernkontexten beschrieben, die im Forschungsprojekt K<sub>3</sub> zum Einsatz kommen. Hierzu werden zunächst Erfolgsfaktoren des netzwerkbasierten Wissensmanagements angeführt und Wirkungsflüsse computervermittelter Kommunikation angesprochen. Auf dieser Grundlage werden Unterstützungselemente für kooperatives E-Learning geschildert, die Bedeutung von Rückmeldungen für Lernprozesse verdeutlicht und Möglichkeiten der Feedbackgestaltung aufgezeigt. K<sub>3</sub> nutzt Feedback sowohl auf Gruppenebene als auch auf der Ebene der einzelnen Lernenden, um positive motivationale und kognitive Effekte zu erzielen. Rückmeldungen werden dabei zu den erzielten Ergebnissen als auch zum Ablauf der Lernprozesse bereitgestellt. Gütekriterien auf Objektebene sind primär qualitative Maße, auf Prozessebene weitgehend strukturquantitativer Art. Diese quantitativen Prozesskennzahlen weisen zwar hinsichtlich ihrer Aussagekraft nur eine sehr geringe Reichweite auf, werden aber in Echtzeit bereitgestellt. Damit sind sie auch ein Mittel, um die verringerte Wahrnehmung in virtuellen Räumen zu kompensieren. Zugleich können sie auch als Analyseinstrument zur Unterstützung der tutoriellen Betreuung genutzt werden. Ergebnisse zu den Feedbackverfahren in K<sub>3</sub> aus verschiedenen Feldstudien verdeutlichen, dass öffentliches evaluatives Feedback, genutzt im Sinne eines fortlaufenden Bewertungssystems, von

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 255-269.

der Mehrzahl der Lernenden als lernerfolgssteigernd und motivationserhöhend eingestuft wird, sich aber auch negativ auswirken und z. B. Angstgefühle hervorrufen kann.

## **I Einleitung**

K<sub>3</sub> steht für Kollaboration, Kommunikation und Kompetenz und ist ein Forschungsprojekt an der Universität Konstanz, das die Umsetzung und Integration kollaborativen Wissensmanagements in der universitären Ausbildung anvisiert<sup>1</sup>. Ziel ist es die Potenziale netzbasierter Wissenskommunikation, Wissensgenerierung und Wissensnutzung für das individuelle und gruppenbezogene Lernen zu nutzen, indem asynchrone Medien – in K<sub>3</sub> Kommunikationsforen – dazu genutzt werden, wechselseitigen Austausch und Kooperation zwischen den Teilnehmern eines Kurses zu befördern [Kuhlen 2002]. Dieser Text zeigt wie in K<sub>3</sub>, neben anderen Unterstützungskomponenten, Feedbackverfahren zur Beförderung netzbasierter Wissenskommunikation genutzt werden. Dazu werden nachfolgend zunächst die grundlegenden Wirkungsflüsse und Erfolgsfaktoren im kooperativen E-Learning genannt und die Bedeutung von Rückmeldungen für derartige Lernszenarien aufgezeigt. Darauf aufbauend wird das ergebnis- und prozessbezogene Feedbackkonzept in K<sub>3</sub> vorgestellt. Dabei wird deutlich, dass Feedback in K<sub>3</sub> nicht nur positive kognitive und motivationale Effekte auf Teilnehmer und Gruppenebene anvisiert, sondern auch über eine Erhöhung der wechselseitigen Wahrnehmung des Handelns der Anderen dahin zielt, technologieinduzierte Defizite der computervermittelten Kommunikation in Foren zu kompensieren.

## **2 Erfolgsfaktoren und Unterstützungselemente netzwerkbauierten Wissensmanagements**

Für die Ausgestaltung kollaborativen Wissensmanagements in Hochschulkursen existieren keine allgemeingültigen Rezepte. Angesichts komplexer Zusammenhänge zwischen den Eigenschaften einzelner Teilnehmer (etwa Medienkompetenz, Vor-

---

<sup>1</sup> K<sub>3</sub> wird an der Universität Konstanz am Lehrstuhl Informationswissenschaft (Prof. Kuhlen) entwickelt. Es handelt sich dabei um ein vom BMBF (DLR PT-NMB+F) im Rahmen des Programms „Innovation und Arbeitsplätze in der Informationsgesellschaft des 21. Jahrhunderts“ in Bezug auf die Fachinformation gefördertes Projekt (Projektnummer: o8C5896). Weitere Informationen unter <http://www.k3forum.net>. Für einen Überblick über K<sub>3</sub> vgl. [Kuhlen et al. 2005].

wissen, Motivation), der Lerngruppen (z. B. Wissensverteilung, Klima, Kohäsion) sowie der Lernumgebung (Curriculare Integration, Didaktisches Design und Technologie) ist es unmittelbar einleuchtend, dass die erzielten Ergebnisse vom Zusammenwirken multipler, interdependenter Wirkungsflüsse abhängig sind [Friedrich & Hesse 2001].

Im Vergleich zur Face-to-Face-Kommunikation weist die Kommunikation in asynchronen Foren erhebliche Mehrwerte auf<sup>2</sup>: Im Kern zum einen die Möglichkeit, Wissenskommunikation auf bislang verschlossene Bereiche auszuweiten, zum anderen ein erheblich erweitertes Rezeptions- und damit Nutzungspotenzial der im virtuell Raum erarbeiteten Kommunikations- und Wissensobjekte. Andererseits kommen für das Lernen in Kleingruppen in asynchronen Umgebungen im Vergleich zum Face-to-Face-Lernen auch erhebliche technikinduzierte Problemfelder zum Tragen, die dazu führen können, dass die erhofften Interaktionsgewinne im Gruppenprozess in asynchronen netzbasierten Lernszenarien u. U. nicht realisiert werden. Hier sind Aspekte reduzierter sozialer Präsenz, sowie eine i. d. R. erhöhte kognitive Belastung anzuführen<sup>3</sup>, die dazu führen dass die soziale Interaktion sowohl auf kognitiver (Grounding) und prozeduraler (Awareness) als auch affektiver Ebene (Soziale Wahrnehmung) behindert wird. Als Problemfelder werden vor allem die beeinträchtigte gegenseitige Wahrnehmung der Teilnehmer und die erschwerte Steuerung des Gruppenprozesses bei der Aufgabenbewältigung gesehen [Jucks et al. 2003].

Insbesondere in Arbeiten die sich dem Forschungsfeld Computer Supported Cooperative/Collaborative Learning (CSCL)<sup>4</sup> zuordnen lassen – vgl. zur Begriffsdiskussion [Hinze 2004] – werden sowohl auf Lernmethodenebene verschiedene instruktionale Unterstützungsmaßnahmen konzipiert als auch auf technischer Ebene eine Vielzahl sogenannter Lerntechnologien<sup>5</sup> entwickelt, die den Zweck verfolgen technologieinduzierte Defizite im kooperativen E-Learning zu kompensie-

---

<sup>2</sup> Mehrwerte herausstreichen, heißt natürlich nicht, die weiterhin bestehenden Vorteile klassischer Kommunikation und herkömmlicher Lernformen geringzuschätzen. Blended Learning trägt zu einer Integration der verschiedenen Kommunikations- und Lernformen bei.

<sup>3</sup> Für eine detaillierte Darstellung vgl. [Griesbaum 2007].

<sup>4</sup> Die Begrifflichkeiten netzwerkbasierter oder kollaborativer bzw. kooperativer Wissensmanagement sowie CSCL und kooperatives E-Learning werden im vorliegenden Text vereinfachend, soweit nicht anders angegeben, synonym behandelt. Zu den wesentlichen Unterschieden und Bedeutungsnuancen der Begriffe vgl. [Griesbaum 2007].

<sup>5</sup> Funktionalitäten, die sich am Lernprozess orientieren oder die spezielle Problemlösungen im Kontext von E-Learning Applikationen darstellen [Niegemann et al. 2004].

ren und die ablaufenden Interaktionsprozesse lernförderlich auszugestalten – vgl. hierzu u. a. die Arbeiten von [Kienle & Herrmann 2004], [Schnurer 2005], [Weinberger 2003].

Neben instruktionalen Unterstützungsmaßnahmen: Kooperationskripten, Beitragstypisierungen, Rollenkonzepten sowie technischen Orientierungs-, Navigationshilfsmitteln und Suchhilfen zur Minderung hypermedia-spezifischen Orientierungsprobleme, setzt K<sub>3</sub> vor allem auch auf konzeptuelle und technologische Möglichkeiten der Feedbackgestaltung, um das netzwerkbasierte Wissensmanagement zu befördern. Dabei stellt sich unmittelbar die Frage, was denn genau unter dem Geben von Feedback verstanden werden kann, ob und inwieweit es kooperative netzbasierte Lernprozesse tatsächlich zu unterstützen vermag und wie es ausgestaltet werden kann. Diese Fragen sind Gegenstand der nachfolgenden, für diesen Text inhaltlich zentralen, Abschnitte.

### **3 Feedback als Instrument zur Unterstützung kooperativer Lernprozesse**

Versteht man unter Feedback jede Art externer Reaktion zu gegebenem Verhalten von Lernenden, so kann dieser Begriff aus einer generischen Perspektive quasi synonym zum kooperativen Lernen betrachtet werden. Sind doch wechselseitiger Austausch und Diskurs, die Kernprozesse kooperativen Lernens, letztlich nichts anderes als Aktionen und darauf folgende Reaktionen der am Lernprozess beteiligten Personen. In didaktischer Perspektive wird Feedback als direkt mitgeteiltes Urteil der Lehrenden über Lernstrategien, Lernfähigkeit und Lernzielerreichung betrachtet [Hargreaves et al. 2000]. Feedback ist damit zunächst ein Maß zur Bestimmung, inwieweit ein Lernziel erreicht wird. Die hohe Bedeutung von Feedback bei Individuen als didaktisches Gestaltungselement in Lernszenarien ist in der Literatur unstrittig<sup>6</sup>. Feedback ist damit ein zentraler Aspekt der Pädagogik bzw. der didaktischen Ausgestaltung von Lernszenarien generell. Die lernförderlichen Aspekte von Feedback lassen sich dabei grundlegend motivationalen und kognitiven Effekten zuordnen, ohne dass diese beiden Aspekte analytisch klar voneinander getrennt werden könnten [Mory 2004]. Um die Wirkungsflüsse (Hinweis-, Lern- und Motivationsfunktion) bei der Feedbackgestaltung in der Praxis besser berücksichtigen zu können, differenzieren [Hargreaves et al. 2000] zwischen evaluativen (Ausdruck der Anerkennung oder Missbilligung der erbrachten Leistung und/oder Belohnungs-

---

<sup>6</sup> [Mory 2004] gibt eine Übersicht über den aktuellen Forschungsstand.

bzw. Sanktionsmechanismen) und deskriptiven (Informationen über die Richtigkeit, Gütegrad von Antworten/Lösungen und das Aufzeigen von Verbesserungsmöglichkeiten) Aspekten des Feedback. [Hey 2001] weist darauf hin, dass Feedback in Gruppenkontexten nicht nur deskriptive und evaluative leistungsbezogene Kriterien beachten, sondern ebenso interpersonelle Faktoren berücksichtigen sollte. Demnach lassen sich vier Dimensionen von Feedback ableiten (Tabelle 1).

	Leistungsdimension	Verhaltensdimension
Individuum	<b>A</b> Individuelles Feedback zur erbrachten Leistung	<b>B</b> Individuelles Feedback über das Verhalten
Gruppe	<b>C</b> Gruppenfeedback zur erbrachten Leistung	<b>D</b> Gruppenfeedback über die Zusammenarbeit

Tabelle 1: Feedback im CSCL in Anlehnung an [Hey 2001], S. 60

Ist sich die Fachwelt über den positiven Effekt von Feedback auf den Lernerfolg weitgehend einig, so ist die Lage hinsichtlich unterschiedlicher Ausgestaltungsformen weniger eindeutig. Es scheint intuitiv naheliegend, dass differenziert ausgestaltete deskriptive Feedbackformen einen hohen Wirkungsgrad aufweisen [Niegemann et al. 2004]. Nach [Jacobs 1998] ist die Effektstärke von Feedbackformen, die mindestens das korrekte Ergebnis beinhalten, höher als bei Feedbackformen, welche nur Aufschluss darüber geben, ob die Antwort richtig war. Allerdings ist die Frage, welche über die Mitteilung des korrekten Ergebnisses hinausgehenden Erklärungen notwendig oder besonders sinnvoll erscheinen, nicht hinreichend erforscht.

Bei der Gestaltung evaluativer Feedbackkategorien ist es sinnvoll, zunächst die grundlegenden motivationalen Orientierungsmuster der Lernenden zu beachten. Nicht alle Lernenden verfolgen primär das Ziel, ihr Wissen zu erweitern bzw. ihre Kompetenzen zu erhöhen (Lernzielorientierung), sondern sind zuerst auf ihre Performanz im sozialen Vergleich mit anderen bedacht [Musch 1999]. Für solche Lernenden ist es oftmals hinreichend, den Anforderungen in einer Weise zu entsprechen, dass sie nicht schlechter als andere eingestuft werden. Lernzielorientierte Lernende sind performanzorientierten Lernenden insbesondere bei schwierigen und arbeitsintensiven Aufgaben überlegen, da sie in solchen Situationen weniger zu Fluchttendenzen, zur Anstrengungsreduktion und zur Abwertung der Aufgabe neigen [Dweck & Legget 1988]. Vor diesem Hintergrund muss Feedback versuchen, eine lernzielorientierte Ausrichtung zu befördern. Dazu sollten anerkennende oder missbilligende Gestaltungskategorien von Feedback nicht den sozialen Vergleich mit Anderen fokussieren, sondern sich auf die früher erzielten Leistungen desselben

Individuums beziehen. Entscheidend für die Wirksamkeit evaluativen Feedbacks ist neben seinem appellativen Charakter vor allem auch das Vorhandensein von Sanktionsmechanismen, denn es ist illusorisch anzunehmen, dass alle Lernenden in hinreichendem Maße intrinsisch motiviert sind bzw. werden könnten [Niegemann et al. 2004]. In kooperativen Szenarien ist der Erfolg der Motivation zudem davon abhängig, dass der Gruppenerfolg an den Erfolg der einzelnen Teilnehmer gekoppelt und die individuelle Belohnung wiederum von der Gruppenbelohnung abhängig ist [Slavin 1993]. Das bedeutet, dass materielle Belohnungsmechanismen sowohl auf individueller als auch Gruppenebene umgesetzt werden sollten. Im kooperativen E-Learning sollten somit bei der Feedbackgestaltung sowohl individuelle als auch gruppenbezogene Rückmeldungen umgesetzt werden.

Aus empirischen Untersuchungen liegen nur wenige Befunde zu Gruppenfeedback vor. U. a. [Hey 2001] untersuchte Effekte von Gruppenfeedback in Unternehmen, unklar ist aber, inwieweit diese Erkenntnisse auf Lernkontexte übertragen werden können. Dennoch lassen sich aus dieser Arbeit Hinweise für die didaktische Ausgestaltung von Gruppenfeedback im CSCL ableiten. Wie z. B., dass neben den ergebnisbezogenen auch zu den prozeduralen Aspekten eine Rückmeldung gegeben werden sollte. Feedback muss also nicht nur die Qualität der Ergebnisse, sondern auch die Qualität des Gruppenprozesses berücksichtigen. Normative Feedbackkriterien können sich dabei an den Gesichtspunkten orientieren, die bei kooperativen Lernprozessen als lernförderlich erachtet werden, so ist insbesondere die Qualität der Interaktion zu beurteilen. Hierzu lassen sich neben eher quantitativen Faktoren, wie etwa der Anzahl oder inhaltlichen Qualität der Beiträge bzw. der Grad der Beteiligung der einzelnen Gruppenmitglieder insbesondere das Ausmaß der Externalisierung von Wissen und das Auftreten integrations- und konfliktorientierter konsensbildender Diskursaktivitäten [Weinberger 2003] als Bewertungsgrundlage des Feedback verwenden. Bei der Gestaltung von Gruppenfeedback ist weiter zu beachten, dass das Feedback kontinuierlich gegeben wird. Fortlaufende Informationen über Quantität und Qualität der erbrachten Leistung und der eingesetzten Methoden im Gruppenprozess sind elementare Faktoren der Fehlererkennung und Problemlösung. [Krause et al. 2003] gehen davon aus, dass auch in Lernkontexten Gruppen Feedback tiefer verarbeiten als Individuen. Zudem kommt in asynchronen computervermittelten Foren eine erhöhte personale Reichweite der Rezeption von Feedback zum Tragen. Je nach dem, wie offen die jeweilige netzbasierte Umgebung gestaltet ist, ist es für die Lernenden nicht nur möglich, das Material und die Ergebnisse anderer Lernender zu rezipieren, sondern auch von den Rückmeldungen zu den Ergebnissen anderer Lernender und Gruppen zu profitieren.



## 4 Feedbackgestaltung in K3

Feedback wird in K3 als Rückmeldung über das Verhalten und die erbrachte Leistung sowohl auf Gruppenebene als auch auf Individualebene im Ablauf der Lernprozesse verstanden. Das Geben von Feedback ist in K3 u. a. Grundlage für ein neues Bewertungssystem, dass die üblicherweise verwendeten punktuellen – z. B. klausurbasierten Bewertungsformen – durch ein fortlaufendes Bewertungssystem ersetzt. Damit wird Feedback in K3 nicht nur deskriptiv, sondern auch als Sanktionsmechanismus genutzt. Abbildung 1 zeigt zur Veranschaulichung das Leistungsbewertungssystem in einem K3-Kurs.

### Leistungsbewertungssystem Kurs Informationsethik 2005

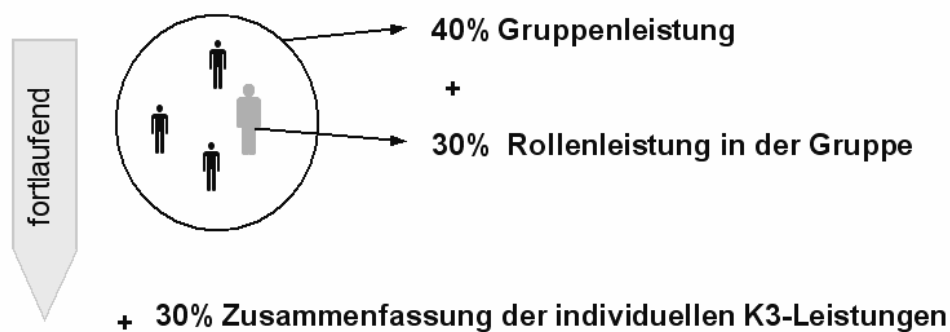


Abbildung 1: Leistungsbewertungssystem im Kurs Informationsethik

K3 stellt ein umfangreiches Instrumentarium zur Verfügung, um Feedback differenziert auszugestalten. Gegenstand des Feedbacks sind auf der einen Seite die im Diskurs erarbeiteten Objekte sowie das Prozessergebnis, auf der anderen Seite wird ergänzend das Verhalten der Teilnehmer im Ablauf der Kollaboration analysiert. K3 stellt hierzu eine Vielzahl von Kriterien zur Verfügung, die sowohl auf Objektebene die Güte der im Diskurs erarbeiteten Diskursobjekte und Referenzen qualitativ und quantitativ messen als auch den Diskurs als Ganzes primär statistisch analysieren. Tabelle 2 zeigt die in K3 standardmäßig vorhandenen Bewertungsobjekte.

K3 ermöglicht es damit auf Objektebene die Qualität individueller und gruppenbezogener Ergebnisse und Prozesse zu erfassen und gegebenenfalls miteinander zu kombinieren bzw. aufeinander zu beziehen.

Die Feedbackkriterien zur Einschätzung der Objekte können in jedem Kurs für alle Bewertungsobjekte frei definiert, um Neue ergänzt und hinsichtlich ihrer Zuordnung zu den verschiedenen Objekten, gemäß den jeweiligen Bedürfnissen des Do-

zenten, angepasst und gewichtet werden. Tabelle 3 zeigt eine kleine Auswahl von bislang in K<sub>3</sub> verwendeten quantitativen und qualitativen Gütekriterien.

<i>Objekttypen</i>	<i>Beschreibung</i>
Kommentar	Ist ein Diskursobjekt auf der Individualebene
Referenz	Ist ein Referenzobjekt (Link, Datei-Upload, Literatur) auf Individualebene
Rolle Moderator	Rollenbewertungen gehen als Individualleistung mit 60% in die Endbewertung ein
Rolle Präsentator	Individualleistung
Rolle Rechercheur	Individualleistung
Rolle Summarizer	Individualleistung
Studentengruppe	Gruppenleistung, geht zu 40% in die Endbewertung ein

*Tabelle 2: Feedback auf Objektebene*

<i>Bewertungskriterien</i>	<i>Beschreibung</i>
Gruppen-Kommentaranzahl	Anzahl der Kommentare der Gruppe (ohne organisationelles)
Absicherungsgrad	Verhältnis eingegebener Wissensseinheiten zur Gesamtanzahl eingegebener Kommentare
Beitragsanzahl	Anzahl der Beiträge eines Mitglieds (Wissenseinheiten, Kommentare).
Einschlägigkeit	Mit Hilfe dieses Kriteriums kann bewertet werden, wie einschlägig ein Beitrag ist.
Ergebnisqualität/Validität	Bewertet die Qualität der Bewältigung der einzelnen Aufgaben und damit des Gesamtarbeitsauftrags. Qualität setzt sich zusammen aus Stringenz/Begründung der Aussagen, Originalität/Selbständigkeit und, Aufgaben-/Arbeitsauftragsrelevanz.

*Tabelle 3: Auswahl quantitativer und qualitativer Gütekriterien*

Die Kriterien sind für die verschiedenen Objekttypen in K<sub>3</sub> von unterschiedlicher Bedeutung. Bei einer Referenz sind vor allem die Einschlägigkeit und der Neuigkeitswert von zentraler Bedeutung, während bei einer Zusammenfassung neben der inhaltlichen Richtigkeit (Validität), vor allem die Vollständigkeit im Sinne der Abdeckung der diskutierten Aspekte, ein zentrales Qualitätskriterium darstellt.

Die Erfahrungen aus mehreren Feldstudien [Griesbaum 2007], [Griesbaum & Ritterberger 2005] haben gezeigt, dass elaboriertes qualitatives Feedback zu individuellen Beiträgen einen sehr hohen Aufwand auf Seiten der Dozenten erfordert. Es ist deshalb sinnvoll die inhaltliche Güteeinstufung auf einige zentrale Beiträge im Diskurs sowie auf das Diskursergebnis zu beschränken. Konkret bedeutet das, Feedback zur inhaltlichen Güte der Gruppenarbeit wird auf Grundlage einer intellektuellen in-

haltsanalytischen Prüfung der Gruppenergebnisse vorgenommen. Zum Geben von Feedback auf der Individualebene werden ergänzend die Leistungen der individuellen Rollen (Moderator, Rechercheur, Zusammenfasser, Präsentator) herangezogen.

Auf der Prozessebene finden primär diskursstatistische Analyseverfahren Anwendung. Diese geben zwar nur in sehr geringem Maße Aufschluss über die Qualität der Kooperation, können aber genutzt werden, um grundlegende Aspekte des Diskursverlaufs zu analysieren und Problemfelder, wie eine Nichtbeteiligung oder eine sehr ungleich verteilte Beteiligung der Lernenden, aufzuspüren, nachzuvollziehen und bei der Feedbackgestaltung zu berücksichtigen. Diskursstatistische Verfahren in K3 bauen grundlegend auf einer Analyse der Zahl der Beiträge im Diskurs auf. Damit lässt sich zunächst der Grad der Beteiligung der Lernenden prüfen. Abbildung 2 zeigt eine solche Analyse.

Statistik der Beiträge bei P2 Was bedeutet informationelle Privatheit bzw. informationelle Autonomie?																			Visualisierung	
Mitglied		Diskursobjekt								Referenzobjekt				Beitragrolle				Statistik		
Name	Rolle	Frage	These	Neues Thema	Ergänzung	Kritik	Resultat	Organisationelles	Andere	Hyperlink	Datei-upload	Lite-ratur	Moderator	Präsentateur	Recher-cheur	Summa-rizer	Summe	%		
		sort	sort	sort	sort	sort	sort	sort	sort	sort	sort	sort	sort	sort	sort	sort	sort	sort		
Lernaktivitäten	M	3	0	5	12	2	1	5	0	8	1	2	21	0	0	0	39	28.06%		
	S	1	5	0	12	3	4	1	0	7	4	0	0	0	2	6	37	26.62%		
	P	0	4	1	11	3	2	1	0	2	2	0	0	9	1	0	26	18.71%		
	HR	0	0	1	15	2	0	1	0	16	0	2	0	0	17	0	37	26.62%		
Summe		4	9	7	50	10	7	8	0	33	7	4	21	9	20	6	139	%		

Abbildung 2: Beitragsstatistik in K3

Das Beispiel zeigt, dass sowohl die einzelnen Objekttypen als auch die Rollenwahrnehmung erfasst werden. Es ist u. a. ersichtlich, dass insgesamt 139 Objekte erarbeitet wurden und der Moderator 39 (rund 28%) Beiträge, davon 21 Moderationsbeiträge, verfasst hat. Diese simple Beitragsstatistik liefert damit einen ersten Überblick über die Art und Anzahl der erarbeiteten Objekte, die Verteilung der Redezeit sowie die Ausübung der Rollenfunktion.

[Semar 2006] entwickelte, aufbauend auf diesen grundlegenden statistischen Daten, verschiedene Kennzahlen, die eine spezifischere Analyse sowohl auf Individual- als auch auf Gruppenebene erlauben. Unterschiedliche Grade auf der Individualebene geben ein detailliertes Bild u. a. darüber wie oft auf Beiträge eines Teilnehmers reagiert wurde (Passiver Reaktionsgrad) und wie häufig ein Teilnehmer auf Beiträge anderer Teilnehmer reagierte (Aktiver Reaktionsgrad). Damit werden auch strukturanalytische Aspekte berücksichtigt und sichtbar. In Anlehnung an das von [Grob et al. 2004] vorgestellte Kennzahlensystem für Learning-Management-Systeme wurden von [Semar et al. 2006] für K3 weitergehende Messzahlen konzipiert, welche zentrale Aspekte der Wissensgenerierungsprozesse auch auf Gruppenebene erfassen. Insbesondere der Kollaborationsgrad liefert ein differenziertes Bild zur Ausprägung des Gesamtdiskurses auf Gruppenebene. Der Kollaborationsgrad

ist dabei ein Quadrupel das aus den Ausprägungen Teilnahmegrad, Synthesegrad, Unabhängigkeitsgrad und Interaktionsgrad gebildet wird (Tabelle 4).

Gruppenkollaborationsgrad $GKG_g$	Ideal	Wertebereich
Synthesegrad $SG_g$	$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$	0 bis 1
Unabhängigkeitsgrad $UG_g$		0 bis 1
Interaktionsgrad $IG_g$		0 bis 1
Teilnahmegrad $TG_g$		0 bis 1

Tabelle 4: Quadrupel „Gruppenkollaborationsgrad“, vgl. [Semar et al. 2006], S. 21

Die einzelnen Messgrößen werden wie folgt gebildet:

- Teilnahmegrad: Auf Individualebene wird der Teilnahmegrad als Verhältnis der Beiträge des jeweiligen Teilnehmers zu allen Beiträgen der Gruppe betrachtet. Auf Gruppenebene wird dieser Wert gemäß [Semar 2006] für alle Teilnehmer zusammengeführt.
- Synthesegrad: Der Synthesegrad beruht nicht auf der statistischen Berechnung von Beitragshäufigkeiten, sondern wird durch die Mitglieder der Gruppen selbst bestimmt, indem sie eine Bewertung zur abschließenden Zusammenfassung abgeben. Der Synthesegrad beruht also auf einer qualitativen Bewertung.
- Unabhängigkeitsgrad: Misst das Vermögen einer Gruppe, autonom, d. h. ohne die korrigierende Hilfe des Lehrenden, ein Ergebnis zu erzielen. Hierzu wird die Zahl der korrigierenden Beiträge des Dozenten ins Verhältnis zur Zahl aller Gruppenbeiträge gesetzt.
- Interaktionsgrad: Der Interaktionsgrad prüft, inwieweit im Diskurs isolierte Beiträge vorliegen. Isolierte Beiträge sind Beiträge, die weder auf einen anderen Beitrag folgen noch selbst einen Kommentar aufweisen. Weiterhin werden Beiträge des Typs Fragen, die keine Antwort aufweisen, als isoliert betrachtet. Der Interaktionsgrad setzt die Zahl der isolierten Beiträge zur Zahl aller Beiträge in Beziehung.

Damit ist es möglich, sowohl die Entwicklung jeder Gruppe als auch die Unterschiede zwischen den verschiedenen Gruppen darzustellen [Semar et al. 2006].

Zur besseren Darstellung der in den Kennzahlendaten enthaltenen Informationen wird in K<sub>3</sub> mit grafischen Darstellungen gearbeitet. Abbildung 3 zeigt einen Vergleich des Gruppenkollaborationsgrades für fünf verschiedene Gruppen.

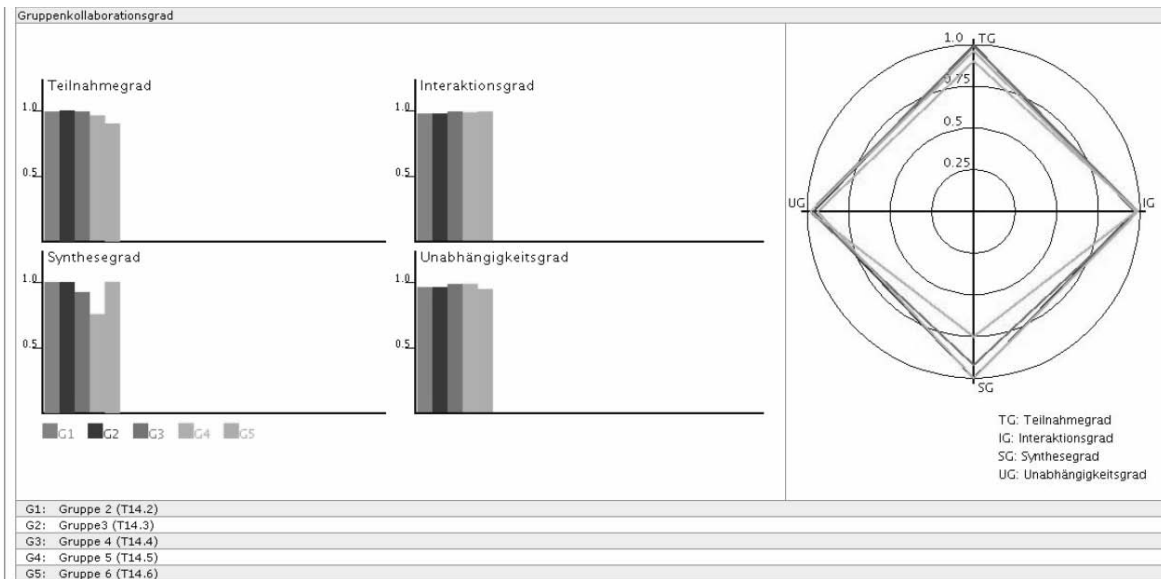


Abbildung 3: Vergleich von Gruppen mit Hilfe des Kollaborationsgrades.

Derartige zwei- oder dreidimensionalen Darstellungen ermöglichen es, Kennzahlen aus verschiedenen Blickwinkeln zu betrachten und erleichtern es sich auf bestimmte Details der Daten fokussieren. Der Einsatz von Visualisierungen befördert also zugleich die Analysemöglichkeiten des Interaktionsablaufs.

Die Darstellung verdeutlicht, dass die Entwicklung von Gütekriterien zur Analyse des Gesamtdiskurses in K<sub>3</sub> weit über die einfache Berechnung elementar beobachtbarer Sachverhalte hinaus geht und mit dem Kollaborationsgrad komplexe quantitative Gütekriterien angewendet werden, die zumindest teilweise auch inhaltsanalytische Bewertungen berücksichtigen. Zu beachten bleibt jedoch, dass wichtige statistische Faktoren, zuvorderst der Umfang auf Textebene, also die Länge der Beiträge, momentan noch nicht erfasst werden. Die Kennzahlen sind ein wichtiges Element der Feedbackgestaltung in K<sub>3</sub>. Der Nutzen der Kennzahlen ist dabei aber nicht auf eine nachlaufende Rückmeldefunktion beschränkt. Vielmehr liegt ein erhebliches Potenzial auch darin, dass der Zugriff auf sie bereits während der laufenden Gruppenarbeit, quasi in Echtzeit, möglich ist. Dadurch werden zugleich die Möglichkeiten der wechselseitigen Wahrnehmung des Handelns der anderen Gruppenmitglieder (Awareness) erhöht. Damit können diese Kennzahlen dazu beitragen, die genannten Problemfelder der verringerten sozialen Präsenz und erschwerten Koordination zu kompensieren. Sie eignen sich auch als Unterstützungselement der tutoriellen Betreuung, da sie im Ablauf indikativ anzudeuten vermögen, ob es sinnvoll bzw. notwendig ist, seitens des Moderators oder des Dozenten intervenierend in den Diskurs einzugreifen [Kuhlen et al. 2006]. Die Kennzahlen zur Analyse des Diskurses sind also ein sehr mächtiges Instrument, dessen Reichwei-

te weit über eine Unterstützungsfunktion zum Geben von nachlaufendem Feedback hinausreicht. Sie sind auch ein Anreizsystem, das die Lernenden so beeinflusst, dass „intrinsische Motivation gefördert wird und daraus eine eigenständige, aktive Teilnahme am kollaborativen Wissenserarbeitungsprozess resultiert“ [Semar et al. 2006].

## **5 Ergebnis und Ausblick**

Zusammenfassend bleibt für die Ausgestaltung des Feedbacks in K<sub>3</sub> festzuhalten, dass eine Vielzahl von Unterstützungskomponenten zur Ausgestaltung von deskriptivem Feedback bereitgestellt wird. Weiterhin bietet K<sub>3</sub> die Möglichkeit, die Erfüllung von Gütekriterien zugleich für das Geben von evaluativem Feedback zu nutzen. Das System gestattet es, die Wichtigkeit der einzelnen Gütekriterien auf einer Skala von 0-10 festzulegen. 0 steht dabei für keine Priorität und 10 für höchste Priorität. Bewertet der Dozent dann den Erfüllungsgrad einzelner Gütekriterien für die Bewertungsobjekte, so errechnet das System automatisch einen Notenvorschlag, der von Lehrenden übernommen oder modifiziert werden kann. Dabei beschränkt sich die Feedbackgestaltung nicht nur auf die Ausgabe berechneter Maßzahlen, sondern kann jederzeit umfangreich paraphrasiert werden und somit in schriftlicher Form sehr elaboriert ausgestaltet werden.

Des Weiteren wird deutlich, dass die in K<sub>3</sub> bislang realisierten strukturquantitativen Kennzahlen sich dazu eignen, grundlegende Aspekte des Diskursverlaufs automatisch zu analysieren, d.h. zunächst Feedback im Sinne von Prozessinformationen bereitzustellen. Durch die damit verbundene Erhöhung der Wahrnehmung des aktiven schreibenden Handelns der Anderen (Awareness) besitzen sie einerseits das Potenzial, einen Beitrag zur Kompensation der verringerten wechselseitigen Wahrnehmung in der virtuellen Umgebung K<sub>3</sub> im Kooperationsprozess selbst zu leisten, als auch andererseits das Geben von Feedback bzw. die tutorielle Betreuung zu erleichtern.

Die Ergebnisse verschiedener K<sub>3</sub>-Feldstudien [Griesbaum 2007], [Griesbaum & Rittberger 2005] deuten an, dass qualitatives Feedback, welches für eine fortlaufende evaluative Bewertung genutzt wird von der Mehrzahl der Lernenden sowohl als motivationssteigernd als auch als lernerfolgserhöhend eingeschätzt wird. Weitergehend zeigt sich aber, dass das im virtuellen Raum erweiterte Rezeptionspotential in Bezug auf das Geben von Feedback nicht zum Tragen kommt bzw. von den Lernenden nicht als Mehrwert empfunden wird. Die K<sub>3</sub> bislang bewusst offen gehaltene Möglichkeit, das Feedback zu anderen Teilnehmern und Gruppen einzusehen,

d.h. die Option vom Feedback zu Anderen zu profitieren, wird von der Mehrzahl der Studierenden weder als motivationserhöhend noch als lernerfolgssteigernd empfunden.

Im Gegenteil, diese Öffentlichkeit des Feedback vermag im Einzelfall auch Angstgefühle zu wecken, welche dazu führen können, dass Diskurse eher gehemmt als gefördert werden [Griesbaum 2007]. Um dieses Angstproblem zu lösen, ist angedacht das Feedbacksystem künftig so zu konfigurieren, dass die Entscheidung des „Öffentlichmachens“ für Andere den Betroffenen selbst überlassen wird. Damit kommen die Vorteile der fortlaufenden Bewertung auf Subjektebene nach wie vor zum Tragen, zugleich werden Angstgefühle bzw. Akzeptanzprobleme gemindert. Der Preis hierfür ist, dass die Subjekte selbst entscheiden, inwieweit das Lernen am Feedback zu Anderen noch möglich ist.

Neben diesem Aspekt stellt sich weitergehend die Frage der Verwendung von Kennzahlen als evaluativen Gütekriterien. Dieser Aspekt wurde im Projekt bislang noch nicht systematisch untersucht. Aus den Kenntnissen der Forschungen zum Geben von individuellem Feedback lässt sich aber befürchten, dass gerade performanzorientierte Lernende, ihr Verhalten gemäß derartiger Kriterien anpassen würden. Die Lernenden etwa nur deshalb auf Beiträge antworten, um einen höheren Reaktionsgrad zu erreichen. In diesem Bereich sind weitere Untersuchungen dringend geboten. Die Frage, ob und inwieweit sie deskriptiv im Sinne prozessanalytische Kennzahlen öffentlich gemacht werden sollen, bleibt ebenso grundsätzlich offen. Zwar ist, wie obenstehend aufgeführt, dadurch eine verbesserte Awareness zu erwarten, u a. die Untersuchung von [Kreijns & Kirschner 2002] verdeutlicht aber, dass die Erhöhung der wechselseitigen Wahrnehmung im Sinne einer Kompensation technologieinduzierter Defizite computervermittelter Kommunikation nicht zwangsläufig mit lernförderlichen Effekten verbunden sein muss.

## **6 Literatur**

- Dweck, C. S.; Legget, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95 Nr. 256, 273.
- Friedrich, H. F; Hesse, F. W. (2001). Partizipation und Interaktion im virtuellen Seminar – ein Vorwort. In: Partizipation und Interaktion im virtuellen Seminar. Friedrich, H. F.; Hesse, F. W. (eds.). Münster, New York, München, Berlin: Waxmann, 7-11.
- Griesbaum, J. (2007). Mehrwerte des Kollaborativen Wissensmanagements in der Hochschullehre – Integration asynchroner netzwerkbasierter Szenarien des CSCL in der Ausbildung der Informationswissenschaft im Rahmen des K3-Projekts. Dissertation Fachbereich Informatik und Informationswissenschaft, Universität Konstanz, Konstanz.

- Griesbaum, J.; Rittberger, M. (2005). A Collaborative Lecture in Information Retrieval for Students at Universities in Germany and Switzerland. In: Proceedings of the World Library and Information Congress: 71st IFLA General Conference and Council. "Libraries – A voyage of discovery", [http://www.ifla.org/IV/ifla71/papers/068e-Griesbaum\\_Ritterberg.pdf](http://www.ifla.org/IV/ifla71/papers/068e-Griesbaum_Ritterberg.pdf) (letzter Zugriff 05.01.2007)
- Grob, H. L.; Bensberg, F.; Dewanto, L.; Düppe, I. (2004). Controlling von Learning Management-Systemen – ein kennzahlorientierter Ansatz. In: Kommen die digitalen Medien an der Hochschule in die Jahre? Carstensen, D.; Barrios, B. (eds.). Münster: Waxmann, 46-56.
- Hargreaves, E.; McCallum, E.; Gipps, C. (2000). Teacher feedback strategies in primary classrooms: new evidence. In: Feedback for learning. Askew, S. (ed.). London: Routledge Falmer, 21-31.
- Hey, A. H. (2001). Feedback und Beurteilung bei selbstregulierter Gruppenarbeit. Berlin: Dissertation.
- Hinze, U. (2004). Computergestütztes kooperatives Lernen. Einführung in Technik Pädagogik und Organisation des CSCL. Münster: Waxmann.
- Jacobs, B. (1998). Aufgaben stellen und Feedback geben, <http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/feedback.pdf> (letzter Zugriff 06.01.2007).
- Jucks, R.; Paechter, M. R.; Tatar, D. G. (2003). Learning and collaboration in online discourses. International Journal of Educational Policy, Research & Practice, 4, 117-146.
- Kienle, A.; Herrmann, T. (2004). Konzepte für die Lerngruppe. In: CSCL-Kompendium. Lehr- und Handbuch zum computerunterstützten kooperativen Lernen. Haake, J.; Schwabe, G.; Wessner, M. (eds.). München: Oldenbourg Verlag, 171-183.
- Krause, U.-M.; Stark, R.; Mandl, H. (2003). Förderung des computerbasierten Wissenserwerbs im Bereich empirischer Forschungsmethoden durch kooperatives Lernen und eine Feedbackmaßnahme.
- Kreijns, K.; Kirschner, P. A. (2002). Group awareness widgets for enhancing social interaction in computer-supported collaborative learning environments. In: Proceedings of the 32nd ASEE/IEEE Frontiers in education conference (session T3E). Budny, D.; Bjedov, G. (eds.). Piscataway, NJ: IEEE.
- Kuhlen, R. (2006). In Richtung Summarizing für Diskurse: In: Information und Sprache. Beiträge zu Informationswissenschaft, Computerlinguistik, Bibliothekswesen und verwandten Fächern Festschrift für Harald H. Zimmermann Herausgegeben von Ilse Harms, Heinz-Dirk Luckhardt und Hans W. Giessen K.G.Saur München, S. 55-74.
- Kuhlen, R. (2002). Vorhabensbeschreibung K3 – Wissensmanagement über kooperative verteilte Formen der Produktion und der Aneignung von Wissen zur Bildung von konzeptueller Informationskompetenz durch Nutzung heterogener Informationsressourcen, <http://www.inf-wiss.uni-konstanz.de/FG/Forschungsprojekte/k3/vorhabensbeschreibung.pdf> (letzter Zugriff 07.01.2007).
- Kuhlen, R.; Griesbaum, J.; Jiang, T.; König, J.; Lenich, A.; Meier, P.; Schütz, T.; Semar, W. (2005). K3 – an E-Learning Forum with Elaborated Discourse Functions for Collaborative Knowledge Management. In: Proceedings of E-Learn 2005 World



- Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education  
October 24-28, 2005 Vancouver BC, Canada. 2981-2988.
- Mory, E. H. (2004). Feedback Research Revisited. In: Handbook of research on educational communications and technology. Jonassen, D. H. (ed.). Mahwah N.J.; London: Lawrence Erlbaum, 745-783.
- Musch, J. (1999). Die Gestaltung von Feedback in computergestützten Lernumgebungen: Modelle und Befunde. Zeitschrift für Pädagogische Psychologie, 13 Nr. 3, 148-160.
- Niegemann, H. M.; Hessel, S.; Hochscheid-Mauel, D.; Aslanski, K.; Deimann, M. und Kreuzberger, G. (2004). Kompendium E-Learning. Berlin, Heidelberg, New York, Hongkong, London, Mailand, Paris, Tokio: Springer Verlag.
- Schnurer, K. (2005). Kooperatives Lernen in virtuell-asynchronen Hochschulseminaren. Eine Prozess-Produkt-Analyse des virtuellen Seminars "Einführung in das Wissensmanagement" auf der Basis von Felddaten. Berlin: Logos Verlag.
- Semar, W. (2006). Kollaborative Leistungsevaluation beim Einsatz von Wissensmanagementsystemen in der Ausbildung. In: Ohly, Peter; Sieglerschmidt, Jörn; Swertz, Christian (Hg.): Wissensorganisation und Verantwortung. Gesellschaftliche, ökonomische und technische Aspekte – Fortschritte der Wissensorganisation 9. Würzburg: Ergon Verlag, 2006, S. 169-177.
- Slavin, R. E. (1993). Kooperatives Lernen und Leistung: Eine empirisch fundierte Theorie. In: Neue Perspektiven der Kooperation – ausgewählte Beiträge der Internationalen Konferenz 1992 über kooperatives Lernen. Huber, G. L. (ed.). Hohengehren: Schneider, 151-170.
- Weinberger, A. (2003). Scripts for Computer-Supported Collaborative Learning. Effects of social and epistemic cooperation scripts on collaborative knowledge construction. München: LMU München: Fakultät für Psychologie und Pädagogik, [http://edoc.ub.uni-muenchen.de/archive/00001120/01/Weinberger\\_Armin.pdf](http://edoc.ub.uni-muenchen.de/archive/00001120/01/Weinberger_Armin.pdf) (letzter Zugriff 07.01.2007).



# Gestaltung von digitalen Lehr-Lern-Szenarien durch offene Innovationsstrategien an Hochschulen\*

*Thomas Sporer, Ulrich Fahrner & Lutz Mauermann*

Institut für Medien und Bildungstechnologie  
Philosophisch-Sozialwissenschaftliche Fakultät  
Universität Augsburg  
Universitätsstraße 10  
D-86135 Augsburg  
*{vorname.nachname}@phil.uni-augsburg.de*

## **I Einleitung**

In den vergangenen Jahren hat sich der Fokus des Medieneinsatzes an Hochschulen von innovativen Projekten einiger Pioniere zur systematischen Integration von neuen Medien in den Regelbetrieb verschoben [vgl. Kerres, 2005]. Es wird daher auf gesamtuniversitäre Medienstrategien gesetzt, bei denen die Alltagstauglichkeit der Innovationen sowie die breite Integration und curriculare Verankerung neuer Medien unter dem Stichwort E-Bologna im Vordergrund stehen. Neuere Entwicklungen im Kontext des Web 2.0 weisen in die Gegenrichtung: Statt formal organisierter Lernprozesse wird besonders das informelle Lernen im Rahmen von selbstorganisierten Praxisgemeinschaften betont, die Lernende als Mitgestalter in die Entwicklung digitaler Lehr-Lernszenarien einbeziehen [Reinmann, Sporer & Vohle, 2007].

Am Beispiel der Universität Augsburg zeigt dieser Beitrag, wie man die Prinzipien offener Innovationen, die auch der Web 2.0-Bewegung zugrunde liegen, im Lehr-Lernbereich einbinden kann: nämlich zum einen bei der Gestaltung von Lehr-Lernszenarien mit digitalen Medien und zum anderen indem man medienversierte Lehrende und Lernende als sog. Lead User in diese Gestaltung mit einbindet [vgl. Chesbrough, Vanhaverbeke & West, 2006, Reichwald & Piller, 2006, von Hippel,

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 271-283.

2005]. Die Darstellung erfolgt in drei Abschnitten: Im ersten Abschnitt wird das Konzept des Augsburger Medienlabors vorgestellt, das Lehrenden und Lernenden einen Raum zur Entwicklung innovativer Prototypen für neue Lehr-Lernszenarien bietet (Kapitel 2). Im zweiten Abschnitt wird die Forschungs- und Entwicklungsarbeit im Medienlabor dargestellt (Kapitel 3). Im dritten Abschnitt wird beschrieben, wie diese Innovationsstrategie im Institut für Medien- und Bildungstechnologie institutionell verankert ist (Kapitel 4). Und im Schlussteil wird reflektiert, wie die didaktischen Innovationen schließlich in den Hochschulalltag diffundieren.

## 2 Medienlabor als Katalysator innovativer Prototypen

Das Medienlabor stellt Studierenden einen „Raum“ bereit, in dem sie eigene Projekte starten und mit neuen Ideen zum Einsatz von Medientechnologien im Studium experimentieren können. Die im Medienlabor vorhandenen Infrastrukturen gestatten es Studierenden, mit geringem organisatorischem Aufwand aus Projektideen funktionierende Prototypen zu entwickeln, welche die Umsetzbarkeit und den Nutzen der Idee demonstrieren. Durch aktives Experimentieren mit dem Prototyp lernen die Studierenden im Idealfall spielerisch, wie man diesen kontinuierlich verbessern kann. Studentische Projektgruppen arbeiten dabei mit Lehrenden zusammen und führen sich die Ergebnisse in Projektgruppen gegenseitig vor. Sie bringen neue Problemlösungen im Entwicklungsteam hervor und planen künftige Weiterentwicklungen. Der gemeinsame Prototyp steht als „Shared Space“ im Mittelpunkt und bildet den Referenzpunkt sowie das Medium einer Konversation mit und über den Prototypen [vgl. Schrage, 1999]. Voraussetzung ist eine Kreativität fördernde Organisationskultur für die Lehrenden und Lernenden, die sich durch folgende Rahmenbedingungen charakterisieren lässt [vgl. Leadbeater, 2006]:

- *Treffpunkte für Lead-User.* Damit sich die Mitglieder einer potenziellen Gruppe von Entwicklern finden, sind Räume erforderlich, um Ideen auszutauschen und gemeinsame Ziele und Intentionen auszubilden.
- *Relevante Infrastrukturen.* Diese Treffpunkte sollten die benötigten IuK-Infrastrukturen bereitstellen und sicherstellen, dass die Initiativen nicht aufgrund mangelnder Ressourcen scheitern.
- *Anreizstrukturen zur Innovation.* Wichtig ist weiterhin, dass es innerhalb der Organisation eine Wertschätzung für innovative Projekte gibt und diese durch Anreizstrukturen auch gezielt gefördert werden.

- *Werkzeuge, Konzepte und Methoden.* Neben den Anreizen zur Innovation sollte die Organisation auch Konzepte, Methoden und Werkzeuge bieten, um die Entwicklungsteams bei der Projektrealisierung zu unterstützen.
- *Testumgebungen für Prototypen.* Um den Erfolg der Prototypen testen zu können, sind Umgebungen bereitzustellen, die zwar den Echtbetrieb im Alltag simulieren, aber die Konsequenzen von Fehlern auffangen.
- *Förderung von Gemeinschaften.* Schließlich fördert die Gestaltung der Treffpunkte als offene Innovationsräume die Entstehung von Communities, die das Engagement und die kreative Entfaltung der Entwickler anregen.

Die im Medienlabor entstehenden Praxisgemeinschaften zeichnen sich dadurch aus, dass Studierende ganzheitlich lernen und durch die Ergebnisse ihrer Projektarbeiten eine echte Wertschöpfung leisten [vgl. Reinmann, Sporer & Vohle, 2006]. Die Ergebnisse dieser Wertschöpfungsleistungen werden, sofern nicht anderweitig vereinbart, nach dem Open Source Prinzip behandelt und können dadurch sowohl im Rahmen der Zielsetzungen des Medienlabors als auch von den an der Entwicklung beteiligten Lehrenden und Lernenden frei genutzt werden. Die Gestaltungsziele des Medienlabors lassen sich dabei in vier Wertschöpfungsbereiche kategorisieren:

- *Ideen und Konzepte.* Dies sind kreative Einfälle für Problemlösungen auf Basis des Erfahrungswissens von Lehrenden und Lernenden sowie Konzepte aus der wissenschaftlichen Community, die strukturiert festgehalten und als Grundlage für Medienentwicklungen dienen können. Diese sollen zum einen praktischen Nutzen haben und zum anderen wissenschaftlich begründen werden können.
- *Praktiken und Prozesse.* Dies sind erfolgreiche Problemlösungen, die sich im Verlauf der Entwicklung eines Prototyps in der Praxis bewährt haben und als „Good-Practice“ beschrieben werden. Bei Praktiken handelt es sich um implizites Wissen, das sich in Handlungen von Personen zeigt. Prozesse sind explizite Beschreibungen einzelner Schritte von Aktivitäten, die bereits dokumentiert sind.
- *Produkte und Services.* Dies sind Ergebnisse der Prototypenentwicklungen, welche aufgrund ihres Erfolgs in den Regelbetrieb übernommen werden sollen. Hierfür werden die Prototypen umfassend getestet und zu Produkten und Services weiterentwickelt. Die Produkte sollen möglichst einfach, zuverlässig und flexibel im Hochschulalltag eingesetzt und bei Bedarf durch Serviceangebote ergänzt werden können.
- *Infrastrukturen und Feedback.* Auf Basis der zuvor entwickelten Produkte und Services werden spezifische Infrastrukturen eingerichtet, welche die langfristigen

Voraussetzungen zum Einsatz der Produkte und Services schaffen. Um die Nachhaltigkeit der erzielten Projektergebnisse zu sichern, ermöglicht ein Feedbacksystem, die Infrastrukturen durch deren Benutzer bedarfsorientiert weiter zu entwickeln.

Die Gestaltungsziele und Handlungsfelder des Medienlabors lassen sich in Form eines Kreislaufprozesses darstellen (vgl. Abbildung 1).



Abbildung 1: Gestaltungsziele und Handlungsfelder des Medienlabors

Die Aktivitäten im Medienlabor umfassen insgesamt vier Handlungsfelder: Den Ausgangspunkt der Arbeit stellt die Forschung zum Lehren und Lernen mit neuen Medien dar. Durch die Beteiligung am fachwissenschaftlichen Diskurs und im Austausch mit Studierenden in der Lehre gehen ständig neue Ideen und Konzepte in das Medienlabor ein. Die Ideen und Konzepte dienen dann in Workshops und Seminaren zur Kompetenzbildung und ermöglichen es, innovationsfreudigen Lehrenden und Lernenden auf Basis der erworbenen Kompetenzen eigene Medienprodukte zu erstellen und neue Praktiken zur Produktion und Nutzung von digitalen Medien zu entwickeln. Erfolgreiche Praktiken werden schließlich als „Good-Practice-Berichte“ dokumentiert und über die Mitarbeiter des Medienlabors weitergegeben. Die Mitarbeiter des Medienlabors helfen wiederum medienaffinen Lehrenden, die beim Einsatz von Medien einen persönlichen Support bevorzugen, bei der Produktion neuer Medieninhalte. Dadurch müssen sich die Lehrenden we-

niger intensiv mit der Medienproduktion befassen, sondern setzen sich gezielt mit der didaktischen Organisation von Lehrinhalten auseinander; damit schaffen sie Lernangebote, die sich durch persönliche Betreuung und prozessbegleitendes Feedback für die Studierenden auszeichnen. Diese Lernangebote werden dann wieder in Lehrveranstaltungen von Studierenden evaluiert und gegebenenfalls weiter verbessert. Die prototypischen Lehr-Lernszenarien, die sich als ausgereift und erfolgreich erweisen, werden letztlich systematisch als Lehr-Lern-Infrastrukturen in den Regelbetrieb von Studiengängen überführt.

Beispielsweise wurde auf diesem Weg die Idee des Medienlabors selbst in ein neues Studienkonzept eingebettet, dass wir „Experience-Centered Technology Studies“ oder kurz „ECTS“ nennen [vgl. Reinmann, Sporer & Vohle, 2007]: *Experience-centered*, weil das erfahrungsgeleitete Lernen in Projekten und Praxisgemeinschaften darin von zentraler Bedeutung ist; *Technology*, weil neuen Technologien im Studienkonzept eine tragende Rolle spielen; und *Studies*, weil wir auf Basis des offiziellen ECTS-Systems eine konsistente Verbindung zwischen dem Lernen in studienbegleitenden Projekten und dem regulären Fachstudium herstellen wollen. Zielsetzung dieses Konzepts ist es, dass Studierende die neuen Technologien als „kognitive Werkzeuge“ erfahren und nutzen lernen [vgl. Jonassen, 1994], indem sie mit geringem Risiko eigene Medienprojekte starten und ein Experimentierfeld haben, um Prototypen zu entwickeln. Diese Prototypen sind als Wissenskonstruktionen der Studierenden zu verstehen [vgl. Papert, 1991], in denen sich das praktische und theoretische Wissen der Studierenden manifestiert: Das Design der Prototypen spiegelt das Wissen der Studierenden wieder [vgl. Perkins, 1987]. In der Gestaltung von Designs liegt für die Lernenden allerdings eine Herausforderung, die von den Lehrenden methodisch angeleitet und unterstützt werden muss [vgl. Schön, 1987].

### 3     **Ansatz des Design-Based Research**

Die Entwicklung der Prototypen im Medienlabor erfolgt methodisch nach den Prinzipien des „Design-based Research“-Ansatzes (DBR). Grundgedanke dieses Ansatzes ist, dass Lehrende und Lernende in natürlichen Umgebungen Design-Experimente machen, durch die sie neue Praktiken und Theorien für das Lehren und Lernen hervorbringen [vgl. Brown, 1992]. Der Erkenntnisprozess ist iterativ und es werden theoriegeleitete Entwicklungsvorhaben mit empirischen Forschungsmethoden kombiniert [DBR Collective, 2003]. Vor dem Hintergrund theoretischer Überlegungen wird eine medienbasierte Lernumgebung entwickelt und

ein Prototyp implementiert. Der Prototyp der Lernumgebung wird im Feld praktisch erprobt und empirisch untersucht – mit quantitativen und/oder qualitativen Methoden. Auf dieser empirischen Basis (der Beobachtungen des Forschers und dem Feedback der Teilnehmer) wird der Prototyp optimiert und die Lernumgebung entsprechend verbessert [vgl. Joseph, 2004]. Neben diesem praktischen Nutzen geht es ebenfalls darum, einen theoretischen Erkenntnisgewinn zu erzielen [vgl. Bell, 2004].

Wie in Abbildung 2 dargestellt, lässt sich der DBR-Ansatz als eine Vier-Felder-Matrix mit zwei Dimensionen vorstellen.

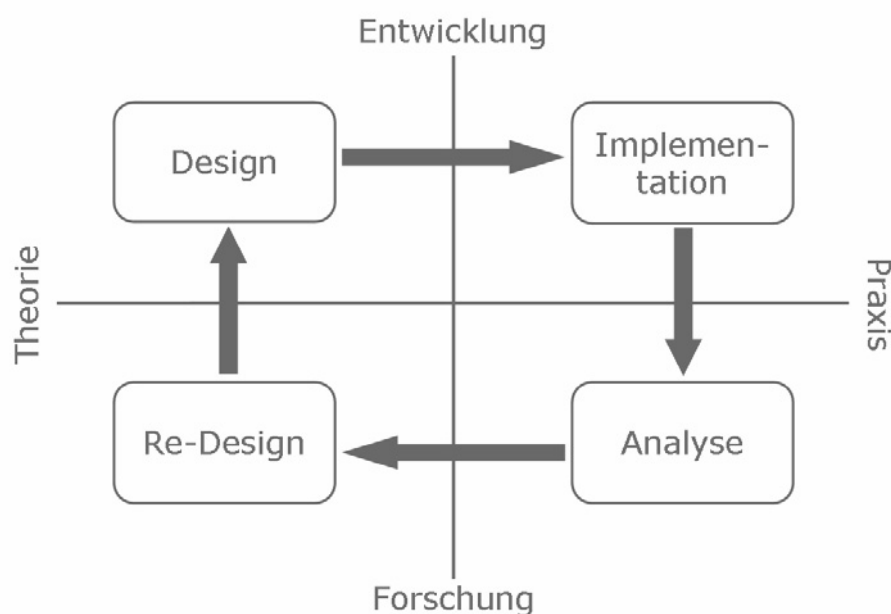


Abbildung 2: Iterativer Innovationsprozess des DBR-Ansatzes

In den vier Feldern wechselt der Fokus des Entwicklungsforschers zwischen den Polen der Theorie- und Praxisdimension sowie zwischen den Polen der Forschungs- und Entwicklungsdimension hin und her, so dass sich folgender Zyklus beschreiben lässt:

- *Design.* Auf der Basis von theoretischen und empirischen Erkenntnissen aus der (Lern-)Forschung wird ein Konzept für einen Prototypen erarbeitet und ein vorläufiges Design für ein neues Lehr-Lern-Szenario entworfen.
- *Implementation.* Der Prototyp wird dann entsprechend diesem theoretischen Design entwickelt und unmittelbar in der Praxis eingesetzt, damit er im Testbetrieb auf die Probe gestellt werden kann.



- *Analyse*. Der Prototyp wird nach der Implementierung in verschiedenen Praxis-kontexten untersucht, wobei die praktische Umsetzung analysiert und die theo-retischen Überlegungen des Designs hinterfragt werden.
- *Re-Design*. Neben den Erkenntnissen aus der Analysephase gehen nun wieder aktuelle Ergebnisse aus der Lernforschung in das Re-Design ein und die Analyse-ergebnisse wirken auf die Theoriebildung zurück.

Beim Durchlauf dieser Phasen des Forschungs- und Entwicklungszyklus werden zwei unterschiedliche Entwicklungslogiken miteinander verbunden [vgl. Schrage, 1999]: Einerseits erfolgt in der Designphase eine bewusste Gestaltung des Lehr-Lernszenarios im Sinne eines Schöpfungsaktes. In der Phase des Re-Designs kommt es andererseits im Sinne des Evolutionsprinzips zur Anpassung des Designs an die Rahmenbedingungen des Echtbetriebs. Beide Logiken lassen sich als zwei komple-mentäre Innovationsprozesse beschreiben [vgl. Gassmann & Enkel, 2006]:

- *Outside-In-Prozess*. Dies meint die Eingliederung von Ideen und Konzepten in die Gestaltung neuer Produkte, Services und Infrastrukturen. Durch Ideen und Konzepten von Außen entsteht ein Design, das im Medienlabor implementiert wird und zu neuen Praktiken und Prozessen führt. Aus der Analyse der Praktiken und Prozesse ergibt sich das Design für neue Produkte und Services, auf deren Basis neue organisationsweite Infrastrukturen für das Lehren und Lernen einge-richtet werden. Wichtig für den *Outside-In-Prozess* ist der Iterationsschritt des Designs.
- *Inside-out-Prozess*. Dies bezieht sich auf die Generierung von neuen Ideen und Konzepten auf der Basis bestehender Produkte, Services und Infrastrukturen, die sich im Umfeld des Medienlabors im Einsatz befinden. Hier können die Infra-strukturen, Produkte und Services, die sich bereits im Regelbetrieb befinden, von den Lehrenden und Lernenden mitgestaltet werden. Dies führt zur Mitgestal-tung von neuen Praktiken, Konzepten und Ideen, wobei sowohl die Perspektive der Lehrenden als Anbieter als auch die der Lernenden als Nutzer berücksichtigt wird. Wichtig für diesen *Inside-Out-Prozess* ist der Iterationsschritt des Re-Designs.

Durch die Rückkoppelung beider Entwicklungslogiken im DBR-Ansatz lassen sich die Chancen zur Realisierung alltagstauglicher Innovationen erhöhen. Insgesamt verlangt dies allerdings von den beteiligten Personen eine breite Palette von Kom-petenzen im Umgang mit persönlichem Wissen und neuen Medien sowie ein hohes Maß an Lernfähigkeit und Veränderungsbereitschaft von Seiten der Organisation [vgl. Reinmann, 2005].

## 4 Institutionalisation des Innovationskonzepts

Das zuvor skizzierte Innovationskonzept ist an der Universität Augsburg durch die Kooperation des Instituts für Medien und Bildungstechnologie (imb) mit dem IT-Servicezentrum (ITS) institutionalisiert. Die Leitidee dieser Kooperation setzt auf eine hybride Strategie, die Elemente offener und geschlossener Innovationsmodelle vereint.

Im Medienlabor des imb wird nach einem offenen Innovationsmodell gehandelt, bei dem Studierende als Entwickler und „Beta-Tester“ agieren und neue Prototypen im Testbetrieb kontinuierlich verbessern. Solchen offenen Organisationsstrukturen lassen sich durch die Eigenschaften selbst-organisierter Communities charakterisieren. Es liegt die Annahme zugrunde, dass der Innovationsprozess dynamisch durch soziale Interaktion inner- und außerhalb der Community entsteht. Die Studierenden und deren Bedürfnisse werden hier als Mitgestalter in den Forschungs- und Entwicklungsprozess einbezogen. Die Mitarbeiter des Medienlabors bilden den festen Kern dieser Organisationsstruktur, während sich die Projekte der Lehrenden und Lernenden an der Peripherie der Organisation in ständigem Wandel befinden [vgl. Brown & Hagel, 2005; Leadbeater, 2006].

- Um bei der Migration erfolgreicher Prototypen in den Echtbetrieb ein solides Dienstleistungsangebot entwickeln zu können, kommt es besonders darauf an, den richtigen Zeitpunkt zu finden, an dem Forschungs- und Entwicklungsprojekte von der ursprünglichen Offenheit zunehmend mehr Struktur annehmen, um die Nachhaltigkeit von Neuentwicklungen zu gewährleisten [vgl. Brown & Duguid, 2001]. Solche Organisationsstrukturen beschäftigen Experten und konzentrieren sich auf die Optimierung der Relation zwischen In- und Output. Die Lehrenden und Lernenden werden hier weniger als Mitgestalter, sondern als Kunden des Wertschöpfungsprozesses verstanden, wobei die Bedürfnisse dieser Kunden durch Marktforschung und ähnliche Methoden erhoben werden [vgl. Leadbeater, 2006].

Die institutionelle Verankerung dieser kombinierten Innovationsstrategie an der Universität Augsburg erzeugt ein Spannungsverhältnis zwischen der kreativen Entwicklung von Prototypen im Medienlabor und den Anforderungen der Überführung von Prototypen in die Strukturen des Regelbetriebs durch das ITS. Dieser Übergang von den informellen Praktiken der Prototypenentwicklung hin zu den formal erfassten Prozessen der Angebote von neuen Produkten, Services und Infrastrukturen im Regelbetrieb, lässt sich schematisch darstellen:

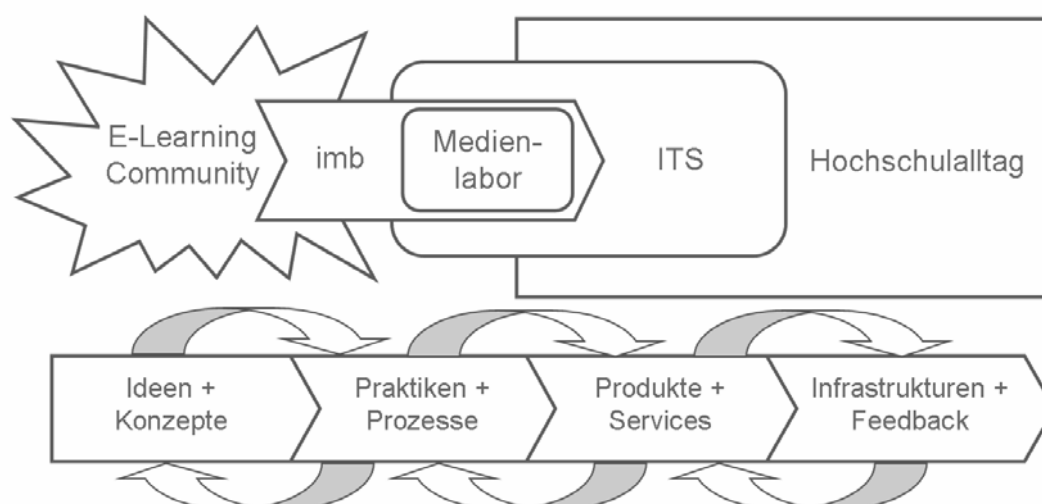


Abbildung 3: Institutionalisierung des Augsburger Innovationsmodells

In Abbildung 3 ist das Augsburger Innovationsmodell als horizontal integrierter Forschungs- und Entwicklungsprozess dargestellt, der aus vier Schritten mit mehreren Mikrozyklen besteht:

- *Innovationsschritt 1.* In der wissenschaftlichen Community zum E-Learning werden von Forschergruppen zahlreiche Ideen, theoretische Modelle und empirische Befunde zusammengetragen sowie Konzepte zum Einsatz von Internet und digitalen Medien zum Lehren und Lernen diskutiert. Das imb greift hier theoretische Konzepte, empirische Befunde und innovative Ideen auf, die dann als Grundlage für die Gestaltung von Lernumgebungen in Forschungs- und Entwicklungsprojekten fungieren.
- *Innovationsschritt 2.* Im Medienlabor des imb werden von Projektgruppen neue Ideen und theoretische Konzepte bei der Entwicklung von Prototypen angewandt sowie deren Anwendungsmöglichkeiten von Lehrenden und Lernenden in Experimenten erprobt. Wenn sich die Prototypen als erfolgreich erweisen, wird das implizite Wissen der Entwickler expliziert, indem die neuen Praktiken in Kooperation mit den ITS-Mitarbeitern erfasst und als Prozesse dokumentiert werden.
- *Innovationsschritt 3.* Die im Medienlabor entwickelten prototypischen Praktiken und Prozesse werden im ITS von Arbeitsgruppen analysiert und in Produkte und Services überführt, die dann von Servicegruppen im Regelbetrieb eingesetzt werden können. Die Entwicklung dieser Produkte und Services erfolgt in bereichsübergreifenden Arbeitsgruppen, die sich aus Mitarbeitern der Projekt- und Servicegruppen des ITS zusammensetzen.

- *Innovationsschritt 4.* Die Produkte und Services werden schließlich durch die Servicegruppen des ITS zu campusweit angebotenen Infrastrukturen des Hochschulalltags. Über das Feedback der Lehrenden und Lernenden im Echtbetrieb fließen neue Ideen und Verbesserungsvorschläge über das ITS in das imb zurück. Durch wissenschaftliche Begleitforschung wird zudem eine Qualitätssicherung bei der Diffusion der Innovationen betrieben. Die Ergebnisse dieser Forschungsaktivitäten gehen letztlich wieder in die wissenschaftliche Fachcommunity ein.

Mit dieser Strategie können die im geschützten Raum des Medienlabors geschaffenen Prototypen aufgegriffen und durch die Kooperation von imb und ITS als systemische Innovationen in der Breite im Hochschulalltag verankert werden [vgl. Chesborough & Teece, 1996]. Die Strategie umfasst dabei alle Phasen der Diffusion von Innovationen [vgl. Rogers, 2003]: In das imb fließen über die Scientific Community aktuelle Erkenntnisse über das Innovationspotenzial digitaler Technologien ein. Bei der Entwicklung von Prototypen im Medienlabor kann man sich zunächst von der Nützlichkeit einer Innovation überzeugen. Liegt dann durch einen funktionierenden Prototypen ein „Proof of Concept“ vor, trifft die Leitung des ITS die notwendigen Maßnahmen dafür, dass die Innovation dauerhaft und campusweit als Produkt bzw. Serviceangebot implementiert wird. Die Akzeptanz der Benutzer gegenüber diesen neu geschaffenen Infrastrukturen entscheidet letztlich darüber, ob sich die Innovation langfristig im Hochschulalltag durchsetzt und in der Breite etabliert.

## 5 Zusammenfassung, Fazit und Ausblick

Dieser Beitrag hat dargestellt, wie an der Universität Augsburg eine offene Innovationsstrategie verfolgt wird, bei der Lehrende und Lernende in die Gestaltung innovativer und zugleich alltagstauglicher Lehr-Lernszenarien einbezogen sind und durch geeignete Forschungs- und Entwicklungsmethoden unterstützt werden kann [vgl. Reinmann, 2006]. Mit dieser Innovationsstrategie sollen zunächst die Lehrenden und Lernenden vom Typus des „Unternehmers“, die Veränderungsprozesse primär als Chance begreifen [vgl. Hagner, 2001], gezielt zur Mitgestaltung von Innovationen durch neue Medien gefördert werden. Damit die Innovation allerdings breitenwirksam in den Hochschulalltag diffundieren kann, muss sich eine kritische Masse an Lehrenden und Lernenden in die mediengestützte Erneuerung der Hochschule einbringen. Über erfolgreiche Projekte der unternehmerischen Pioniere hinaus sollte eine Strategie daher auch eine institutionelle Förderung und öffentlichkeitswirksame Anerkennung für die Innovationstypen der „Karriereorientierten“

und der „Risikovermeider“ bieten [vgl. Hagner, 2001]. Eine solche Strategie zielt darauf ab, die eher karriere-orientierten Lehrenden und Lernenden durch Anreizsysteme zu aktivieren und die eher zögerlichen Lehrenden und Lernenden durch Supportstrukturen zu unterstützen [vgl. Seufert & Euler, 2005]. Mit diesen aktiven Maßnahmen zur Förderung von Innovationen im Hochschulalltag lassen sich die Gruppen der Innovatoren, der frühen Adaptoren sowie der frühen Mehrheit erreichen [vgl. Rogers, 2003]. Um bei den Gruppen der späten Mehrheit und den NachzöglerInnen ebenfalls eine Veränderungsbereitschaft zu erreichen, empfehlen sich Maßnahmen eines „Minimal Change“ [vgl. Kerres, 2005].

Da an der mittelgroßen Campusuniversität Augsburg nur begrenzte personelle und finanzielle Ressourcen zur aktiven Erneuerung der Hochschule durch neue Medien zur Verfügung stehen, konzentriert sich die Medienstrategie auf die Förderung der Aktivitäten von Inventoren und veränderungsbereiten Adaptoren von didaktischen Innovationen. Die in diesem Beitrag vorgestellte Innovationsstrategie umfasst damit folgende Elemente:

- *Strategieelement 1.* Förderung von innovativen Projekten auf „Low-Budget-Basis“ durch die Einbindung von Lehrenden und Studierenden in die Entwicklung von Prototypen für mediengestützte Lehr-Lernszenarien (vgl. Gestaltungsziele des Medienlabors in Kapitel 2).
- *Strategieelement 2.* Gezielte Unterstützung der Entwicklungsaktivitäten von Lehrenden und Studierenden durch das Institut für Medien- und Bildungstechnologien (imb) mit Schwerpunkt auf der E-Kompetenzbildung (vgl. Handlungsfelder des Medienlabors in Kapitel 2)
- *Strategieelement 3.* Unmittelbare Einbindung von studentischen Medienprojekten aus dem Studiengang „Medien und Kommunikation“ im Sinne eines unternehmerischen Lernansatzes (vgl. ECTS-Konzept des Begleitstudiums „Problemlösekompetenz“ in Kapitel 2).
- *Strategieelement 4.* Stetige Verbesserung der Prototypen und Modelle durch die Rückkopplung von Nutzerinteresse bei der Medienentwicklung und Erkenntnisinteresse bei der wissenschaftlichen Begleitforschung (vgl. „Design-Based Research“ in Kapitel 3).
- *Strategieelement 5.* Systematische Integration von erfolgreichen Prototypen in den Hochschulalltag durch die Supportstrukturen des IT-Servicezentrums (ITS) als Koordinationsstelle für IuK-Dienstleistungen an der Uni Augsburg (vgl. Integriertes Innovationsmodell in Kapitel 4)

Diese kombinierte Strategie strebt eine proaktive Gestaltung der Lernkultur der Hochschule sowie eine Reform der Studienorganisation im Kontext des Bologna-Prozesses an [vgl. Reinmann, Sporer & Vohle, 2007]. Neben dieser Reformstrategie auf Basis der kollaborativen Gestaltung der Hochschule von Lehrenden und Lernenden wird im Hinblick auf bestehende Studienangebote auch eine Flexibilisierungs- und Professionalisierungsstrategie verfolgt [vgl. Seufert & Euler, 2005]. Bei der Umsetzung dieser Ziele setzen wir nicht auf „Push-Programme“, sondern orientieren uns am Modell einer „Pull-Plattform“, welche hilft, verteilte Ressourcen zu mobilisieren und zu integrieren [vgl. Brown & Hagel, 2005].

## 6 Literatur

- [Bell, 2004] Bell, P. (2004). On the Theoretical Breadth of Design-Based Research in Education. *Educational Psychologist*, 39(4) 243-253.
- [Brown, 1992] Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *Journal of the Learning Sciences*, 2(2), 141-178.
- [Brown & Hagel, 2005] Brown, J.S. & Hagel, J. (2005). From Push to Pull: The Next Frontier of Innovation. Some companies are learning how to take a more creative approach to mobilizing resources. *McKinsey Quarterly*, 2005 Number 3
- [Brown & Duguid, 2001] Brown, J.S. & Duguid, P. (2001). Creativity versus structure: A useful tension. *MIT Sloan School of Management Review* (Summer), 93-94.
- [Chesbrough, Vanhaverbeke & West, 2006] Chesbrough, H.W., Vanhaverbeke, W. & West, J. (2006): *Open Innovation. Researching an New Paradigm*. Oxford: Oxford University Press.
- [Chesbrough & Teece, 1996] Chesbrough, H.W. & Teece, D.J. (1996): Innovation richtig organisieren – aber ist virtuell auch virtuos? In: *Harvard Business Manager*, Vol. 18/3, S. 63-70.
- [Collins 1999] Collins, A. (1999). The Changing Infrastructure of Education Research. In E. C. Lagemann & L. S. Shulman (Eds.), *Issues in Education Research: Problems and Possibilities* (pp. 289-298). San Francisco: Jossey-Bass.
- [DBR Collective, 2003] Design-Based Research Collective (2003). *Emerging Paradigm for Educational Inquiry*. 8. Online at <http://www.aera.net/pubs/er/pdf/>
- [Gassmann & Enkel, 2006] Gassmann, O. & Enkel, E. (2006): Open Innovation. Die Öffnung des Innovationsprozesses erhöht das Innovationspotential, in: *zfo*, 3/2006 (75. Jg.), S. 132-138
- [Hagner, 2001] Hagner, P.R. (2001): Interesting practices and best systems in faculty engagement and support. Vortrag auf der NLII Focus Session, Seattle.
- [Jonassen, 1994] Jonassen, D.H. (1994). Technology as cognitive tools: learners as designers. *ITForum*, paper #1. Online publications edited by Gene Wilkinson,

- Department of Instructional Technology, University of Georgia. Available online:  
<http://itech1.coe.uga.edu/itforum/paper1/paper1.html>
- [Joseph, 2004] Joseph, D. (2004). The Practice of Design-Based Research: Uncovering the Interplay Between Design, Research, and the Real-World Context. *Educational Psychologist*, 39 n4 p235-242.
- [Kerres, 2005] Kerres, M. (2005). Strategieentwicklung für die nachhaltige Implementation neuer Medien in der Hochschule. In Th. Pfeffer, A. Sindler, A. Pellert & M. Kopp (Hrsg.) *Handbuch Organisationsentwicklung: Neue Medien in der Lehre – Dimensionen, Instrumente, Positionen* (S. 147-162). Münster: Waxmann.
- [Leadbeater, 2006] Leadbeater, C. (2006) *The user innovation revolution: how business can unlock the value of customers' ideas*. London: NCC.
- [Papert, 1991] Papert, S. (1991). *Situating Construction. Constructionism*, I. Harel & S. Papert (eds.) Norwood, NJ: Ablex Publishing.
- [Perkins, 1986] Perkins, D.N. (1986). *Knowledge as design*. Hillsdale, NJ: Lawrence Erlbaum.
- [Reichwald & Piller, 2006] Reichwald, R. & Piller, F. (2006). *Interaktive Wertschöpfung. Open Innovation, Individualisierung und neue Formen der Arbeitsteilung*. Wiesbaden: Gabler Verlag.
- [Reinmann, 2005] Reinmann, G. (2005). Lernort Universität? E-Learning im Schnittfeld von Strategie und Kultur. *Zeitschrift für Hochschuldidaktik*, 6, 66-84. Internet:  
[http://www.zfhe.at/resources/downloads/FHD\\_06\\_009\\_REINMANN\\_Lernort\\_Universitaet\\_1000605.pdf](http://www.zfhe.at/resources/downloads/FHD_06_009_REINMANN_Lernort_Universitaet_1000605.pdf)
- [Reinmann, 2006] Reinmann, G. (2006). Nur „Forschung danach“? Vom faktischen und potentiellen Beitrag der Forschung zu alltagstauglichen Innovationen beim E-Learning. *Arbeitsbericht Nr. 14 der Medienpädagogik an der Universität Augsburg*.
- [Reinmann, Sporer & Vohle, 2007] Reinmann, G., Sporer, T. & Vohle, F. (in Druck). *Bologna und Web 2.0: Wie zusammenbringen, was nicht zusammenpasst?* In Kerres, M. & Keil, R. (Hrsg.) *eUniversity – Update Bologna. Education Quality Forum*. Bd. 3, Münster: Waxmann.
- [Rogers, 2003] Rogers, E.M. (2003). *Diffusion of Innovation*, Fifth Edition. New York, NY: Free Press.
- [Schön, 1987] Schön, D. A. (1987). *Educating the reflective practitioner. Toward a new design for teaching and learning*. San Francisco: Jossey-Bass.
- [Schrage, 1999] Schrage, M. (1999). *Serious Play. How the World's Best Companies Simulate to Innovate*. Boston: Harvard Business School Press.
- [Seufert & Euler, 2005] Seufert, S. & Euler, D. (2005): *Nachhaltigkeit von eLearning-Innovationen: Fallstudien zu Implementierungsstrategien von eLearning als Innovationen an Hochschulen (SCIL-Arbeitsbericht 4)*: St.Gallen: SCIL. Internet:  
<http://www.scil.ch/publications/docs/2005-01-seufert-euler-nachhaltigkeitelarning.pdf> (Stand: 01.09.2005).
- [von Hippel, 2005] von Hippel, E. (2005): *Democratizing Innovation*. Boston, MA: MIT Press.





# Extracting Informative Content Units in Text Documents\*

## Using Topic Chains for Conceptual Document Representation

*Jürgen Reischer*

Information Science  
University of Regensburg  
93040 Regensburg  
*juergen.reischer@sprachlit.uni-regensburg.de*

### Abstract

The notion of semantic and thematic informativeness of text is explored in theory and practice. The IVal system is described which implements a procedure for conceptual text analysis and representation based on thematic chains. Possible applications for semantic text processing including conceptual indexing and passage extraction are presented and discussed.

### I Introduction

The notion of *informativeness* has not gained much attention in the literature, although informativeness is the primary quality an information seeking person certainly expects of documents or passages retrieved. Contrary to the frequently discussed and defined concept of information,<sup>1</sup> the notion of informativeness lacks an explication which also impedes its application in computational language processing. For example, in order to detect and select informative content units in text documents, we need both a conception of informativeness and a procedure for automatic extraction of informative passages. In the following sections, therefore,

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10. International Symposium for Information Science. Constance: UVK, 285-302.

<sup>1</sup> An extensive overview and discussion about ordinary and scientific concepts of information can be found in [Reischer 2006a].

we will address theoretical and practical aspects of informativeness in view of its explication and automatic extraction.

## 2 The Concept of Informativeness

We will approach the term 'informativeness' from both the perspective of everyday as well as scientific language. After that, we will give an explication of informativeness as used in this context and discuss related notions.

### 2.1 Informativeness in Everyday Language

The term 'informativeness', as derived from 'informative', may be understood from several perspectives if we consider its everyday meaning: on the one hand, an expression (sentence, text) may be informative *for itself* without reference to a certain interest or information need, e. g. the sentence "I am here now and doing something" is less informative than "J. R. works on his paper in Regensburg on 24/12/2006 at 15.00 o'clock"; on the other hand, a text (passage or document) may be informative *about* a certain topic or *relative to* an information need of a user (as expressed by search terms). Another distinction of 'informative(ness)' concerns the semiotic levels of syntax, semantics, and pragmatics. Primarily, informativeness is related to the *semantic* level of text in the sense of concepts or propositions denoted by signs. Syntactic informativeness may be understood, for example, as the information content of signs derived from their probability of appearance [cf. Shannon 1948], where 'sign' just means a symbolic *form* (irrespective of its meaning). Pragmatic informativeness, finally, can be interpreted as such instructive and/or enlightening semantic contents that immediately promote or prohibit action.

Besides these qualitative distinctions, informativeness is also a quantifiable concept, i.e. text may be *more* or *less* informative. On the one hand, this may simply mean that it contains more or less informational units measured absolutely or relatively to its length. On the other hand, it may be interpreted as the degree of specificity or preciseness of concepts and propositions conveyed by the text. In the first case, we have more pieces or bits of information; in the latter case, we simply gain more informational content.

### 2.2 Informativeness in Scientific Context

Departing from the ordinary understanding of 'informativeness', we will have a closer look at some uses of this notion in scientific contexts. Like many other terms,

'informativeness' is used ambiguously here; I will just give some examples. In Tague-Sutcliffe, informativeness simply means the *amount of information* conveyed by the documents or records provided by an information service [Tague-Sutcliffe 1995]. Qualitatively, information is to be understood here as conceptual content in accord with Fox' definition of information as propositions [cf. Fox 1983]. In the process of abstracting, informative summaries cover the *most important concepts*, as contained in corresponding passages of a text [cf. e. g. Mani & Maybury 1999]. In text linguistics, informativeness means the extent to which textual elements (words, sentences) are *expected* versus *unexpected* or *known* versus *unknown* [Beaugrande & Dressler 1981]. This explication seems like a semantic counterpart to Shannon's (syntactic) surprise value of signs. Rosch and colleagues consider informativeness of concepts from the perspective of basic level categories and their super- and subordinate concepts. In this regard, basic level concepts have optimal informativeness with respect to the *number of defining attributes* and the *number of attributes differentiating it from neighbour concepts*, thus minimizing the cognitive load for information processing [cf. Rosch & Mervis 1975; Rosch & al. 1976; Rosch 1978]. The author himself provides a measure of informativeness for concepts based on the assumption that the (vertical) position of a concept in a conceptual hierarchy (e. g. as realized in WordNet) is an indicator of its *semantic specificity* [Reischer 2006b]. The more specific a concept is, the lower it is positioned in the hierarchy and the more distinctive features it must carry, thus being more informative (as in 'This is a *collie*' versus 'This is a *dog*').

In summary, this short extract from the literature reveals a quite inhomogeneous usage of the term 'informativeness' which prevents a synopsis of the underlying conceptions. Nevertheless, the notion of informativeness primarily applies to semantic units like concepts as conveyed by signs (word, sentence, text), which may be more or less informative. The amount of information transmitted by linguistic expressions (on a certain topic) may be called the semantic-thematic or conceptual information content of a text. Because concepts are regarded as purely semantic units, they cannot be accessed immediately from the textual surface which consists merely of a sequence of *terms* in the sense of *morphosyntactical* units. Terms are assigned one or more concepts as their meaning (signification), so that these concepts can be accessed indirectly via the terms expressing them. Because many terms signify more than one concept, i.e. they are ambiguous, they have to be disambiguated to select the correct semantic information the term conveys in a certain (co)text.

In the next sections, we will give an explication of informativeness in the sense of semantic and thematic (= conceptual) information content.

### 2.3 An Explication of Informativeness

Informativeness could be understood as the potential of a text or a text passage of a document to satisfy a user's information need (better or worse). This may either be understood as the general or absolute capacity of a text to be informative as such (for any recipient in any context), or as the specific or relative capacity of a text to be informative about a certain topic or with respect to the information need of a user (as reflected by some search terms in either case). A text is more or less informative if it has the potential to convey a certain amount of semantic or thematic information (on a certain topic), i.e. it has a lower or higher information content. Necessarily the question raises, what properties of a text account for its informativeness.

Basically, semantic information is carried by single concepts or complex propositions constituted of concepts by forming predicate-argument structures. Although it would be highly desirable to compute propositional information contents automatically, we must restrict ourselves to the information contained in single concepts. With respect to computational aspects this can be realized with much less effort than completely parsing text for its syntax and semantics. Furthermore, conceptions of propositional information content in the literature are hardly translatable into adequate automatic procedures. In the following sections, therefore, we will go into details about the notions of computable semantic and thematic information content which can be used for practical applications.

#### 2.3.1 *Semantic Information Content*

The notion of semantic information content may be tackled from two different perspectives: We may either look at properties of the concepts themselves or at the distribution of concepts within a text. In the first view, information content may be related to semantic specificity (as opposed to genericity) or semantic frequency of occurrence (of concepts); in the latter view, semantic information can be related to semantic density.

The notion of semantic specificity versus genericity of concepts is based on the idea that a concept contains the more information the less other concepts it classifies (and vice versa). For example, the concept ENTITY classifies all other concepts, because everything is an entity; in contrast, the concepts SEPTEMBER-11-2001 or ALBERT-EINSTEIN classify no further concepts, because they are informationally as discriminative as possible denoting only one specific thing in the world. On the linguistic level, a hyponym (subordinate) of a term carries more information than its hyperonym (superordinate). For instance, 'collie' – or better: its meaning – is more

specific than 'dog', and 'Lassie' is more specific than 'collie'. If I am said that X is a collie, I gain more knowledge than if I hear that X is an animal. In order to automatically compute information contents for concepts in the sense of semantic specificity, we need a conceptual hierarchy with hyperonym-hyponym relations between these concepts. One such hierarchy is available with WordNet (or GermaNet or any other net), where nouns and verbs are organised into hierarchical trees from the most generic to the most specific concepts. An algorithm for computing information values on the basis of the vertical position of a concept in the hierarchy can be found in [Reischer 2006b].

In an alternative measure of semantic information, the frequency account of Shannon's information content (Shannon 1948) is combined with the above approach based on conceptual hierarchies. The basic assumption behind this approach is twofold: firstly, the observation that frequently recurring entities carry less information than rarely occurring entities (compare the more or less probable symbols of an alphabet in Shannon's theory);<sup>2</sup> secondly, the occurrence of a semantically more specific concept X necessarily entails the *conceptual* cooccurrence of a more general concept Y superordinated to X.<sup>3</sup> That is to say, if we talk of collies, we also talk of dogs; if we think of Lassie, we also think of collies. Consequently, all occurrences of concepts  $X_Y$  subordinated to Y contribute to the *semantic* occurrence of Y, because we cannot think of  $X_Y$  without thinking Y. The semantic information content of Y, then, is in analogy to Shannon, the binary logarithm of the frequency  $F(Y)$  of a concept in question *including the frequencies of all its subordinate concepts  $X_Y$*  in relation to the frequency of the upmost root concept  $F(R)$  including the summed-up frequencies of all its subordinates  $X_R$ :  $I(Y) = -\log_2 (F(Y) / F(R))$ , where both numerator and denominator must be greater than 0. For normalization, we must divide by the maximum information value possible (i.e. if a concept has frequency 1):  $I_N(Y) = [\log_2 (F(Y) / F(R))] / [\log_2 (1 / F(R))]$ . For  $F(Y) = 1$ , we get  $I_N(Y) = 1$  (maximum information content), for  $F(Y) = F(R)$ , we get 0 (minimum information content). A low information content means that a concept Y occurs frequently, both by direct activation via an appropriate term (e.g. 'thing' etc.) and by being activated indirectly via all subordinated concepts and their respective terms (e.g. 'dog', 'collie', 'Lassie' etc.). The higher the frequency of a concept Y the

<sup>2</sup> In this approach, we are dealing with frequencies of *concepts* as semantic units, not with terms as symbols in the sense of morphosyntactic units.

<sup>3</sup> See Resnik (1995) for a comparable approach in semantic similarity measurement. The *concept* frequencies required can be derived from the WordNet database where the frequencies of terms *in certain senses* (i.e. concepts) are available.

more subordinates  $X_Y$  it typically has contributing to the frequency of  $Y$ , i.e. the more upward in the conceptual hierarchy and the more unspecific it must be. Necessarily, such a concept occurs in many communicative contexts and, therefore, must be indiscriminative with respect to the information it carries (because it has to be compatible with many other concepts in order to be semantically combinable with them).<sup>4</sup>

The notion of semantic density or concentration may be explicated simply by the ratio of the number  $N_C$  of content words to the number of function words  $N_F$  in a text or passage with  $N_C + N_F$  words:  $D = N_C / (N_F + N_C)$ . The more content elements a text contains, the less uninformative (formative) units it can have (at equal length). The ratio is 1 if all items in the passage are content words, as in telegrams like 'Coming Sunday morning John'; the ratio is 0 in sentences with only function words like 'I am here now'. Content words are typically nouns, verbs, adjectives, and many adverbs; function words are basically prepositions, conjunctions, pronouns, and determiners.<sup>5</sup> This quite simple measure of semantic informativeness is based on the idea that a text (passage) cannot convey more semantic information than is expressed by content words.

### 2.3.2 *Thematic Information Content*

Thematic informativeness is basically concerned with the topic structure of a text. A text is thematically more or less informative, if it conveys one or more topics in a more or less cohesive/coherent and concentrated (dense) way. Cohesion and coherence concern the syntactic and semantic connectedness of textual units (terms and concepts), respectively, to form a thematic whole [cf. Beaugrande & Dressler 1981]. Concentration means the density of these units with respect to a certain length of a textual passage.

The notions of cohesion and coherence are related to the syntactic and semantic means a text (writer) uses to form an interconnected sequence of words and sentences, concepts and propositions [cf. e. g. Halliday & Hasan 1976; Hoey 1991; Morris & Hirst 1991]. Thematic convergence of these textual units guarantees that the text is not an accidental, senseless concatenation of words and sentences but forms a consistent whole containing and conveying one or more topics. On the co-

---

<sup>4</sup> The content of this paragraph is taken from [Reischer 2007b].

<sup>5</sup> It is certainly debatable whether all prepositions are really function words, e. g. those derived from verbs like 'corresponding to' and 'according to'. Another point is the question in how far pronouns are not informative: if they are treated like the nouns they stand for, they *are* of course informative.

hesion level, this is accomplished by at least two syntactic means: pronouns refer back to thematic units already introduced (functioning like definite noun phrases); collocations (in the two senses of lexicalised multi-word expressions like idioms and otherwise contextually co-occurring terms) connect isolated words to meaningful clusters. On the coherence level, two phenomena connect concepts and propositions: semantic coreference and thematic associations between single concepts constitute chains of conceptually related elements (e. g. car – automobile, car – steering-wheel), conjunctions between propositions relate one complex content to another. The more cohesive and coherent units a text contains, the more thematically convergent it is assumed to be due to its formal and conceptual interwovenness. Computationally, this may be realized by a procedure assembling conceptually related elements to thematic chains: the length and strength of such chains are an indicator for the thematic information content (topicality) of the text (see section 3.1 below for further elaboration).

Thematic concentration or density means that some conceptual units of a text are such concepts constituting a topic of the text sustaining many thematic (conceptual) relations to other units in the same passage or text. Furthermore, the average thematic distance between all concepts is comparatively low so that many or even all conceptual units are strongly interconnected to a thematic network leaving only few concepts isolated. For example, the relatedness between concepts like INTERNET, FIREFOX, and BROWSER is very strong, whereas the connection between the concepts INTERNET, CELLPHONE, and CABLE is somewhat looser; in contrast, the concepts INTERNET, LASSIE, and JUSTICE constitute no intrinsic thematic relation. Another indicator of thematic concentration may be the type-token distribution of the text's concepts: if a text has few concepts with high frequency and many concepts with low frequency (relative to the total number of concepts), it is thematically focussed; if we have a text with a more homogenous type-token profile, the topic seems more diffuse or distributed. However, in both cases we have to consider 'thematic noise', i.e. the relative number of thematically singular and unconnected concepts being isolated from all others (see below).

#### **2.4 Some Annotations to the Notion of Informativeness**

Informativeness, understood as semantic or thematic information content, is intrinsically related to the notions of text comprehensibility and aboutness or relevance. Some details may be indicated in the following.

With respect to semantic understandability, as opposed to the linguistic surface phenomena of legibility and readability, a text is the less comprehensible the more

(new) information it codes in a certain passage of text, be it the concentration of content units (propositional density) per formal unit, conceptual specificity (optimal basic concept level), or the explicit presence of sentence connectives [cf. e. g. Rosch & al. 1976; Kintsch & Vipond 1979]. That is to say, if a text is hard to understand, the probability of misunderstanding or even non-understanding increases so that the text will not be as informative for the recipient as potentially possible. Under this view, we have to discriminate between objective and subjective informativeness: the amount of information coded in the text is its theoretical *potential* that can be conveyed, the amount of information decoded by the recipient is the *actual* information arriving at the user. Consequently, if we want to quantify subjective information content, we have to consider general cognitive factors of comprehensibility (e. g. frequent and/or basic level concepts are even understandable by children) as well as specific factors of the recipient (e. g. his preknowledge of a domain). We are primarily concerned here with objective informativeness because subjective informativeness necessarily depends on the latter; i.e. the recipient cannot extract more semantopragmatic information than is semantically coded in the text.

The concept of relevance and its different interpretations in information science was summarized and discussed in [Mizzaro 1997]. As a common ground, relevance is a relational concept: something is relevant *for* something else. In contrast, informativeness can be absolute or relative: something is informative (as such) versus informative for/about something else. In the latter case, informativeness-for/about includes relevancy: a document is informative about a certain topic/question or for the information need of a user, if it is at least (thematically) relevant to the latter *and* if it is understandable and qualitatively well-built (e. g. orthographically well-formed, thematically coherent). Thus, a text document rated as relevant is actually not relevant for the user if it is completely incomprehensible or obviously lacks quality control. The user wants informative documents, not just relevant ones. This should be considered at least in passage or document ranking.

### **3 Realization of Informativeness Evaluation**

The automatic detection and selection of informative contents in text documents can serve two different goals: firstly, in the text-oriented view, we want to extract those contents which are most 'representative' for the content of a text, i.e. an index of concepts (and respective terms representing them) or a list of passages being most informative about the topic(s) of the entire document (indexing and summarizing); secondly, in the user-oriented view, we want to extract those sections of the text that



match best with the required contents expressed by one or more search terms representing the concepts desired (sentence/section retrieval). In addition to the detection and selection of the most informative passages with respect to thematic content, we may further select those passages from the list of candidates with the highest/lowest semantic information contents (ranking of passages, e. g. for comprehensibility, conciseness, detailedness, etc.).

The two scenarios presented above unify several information processing tasks to one general conception of generally detecting and selecting informative contents in text documents: summarizing is viewed as the extraction of the most informative passages based on one or more concepts being thematically most representative for the entire text (retrieval by the most important concepts of the text taken from its concept index); passage retrieval is conceptionally based on the same mechanism replacing internal by external, user-defined search terms and their associated concepts. Thus, we can efficiently implement summarizing (as extraction of sections) and passage retrieval with the same procedure to be described in the next section.

### **3.1 Implementation**

In the following subsections, the first implementation of a general procedure for semantic informativeness extraction on the basis of the WordNet and IVal systems and their applications will be described in more detail.

#### *3.1.1 The WordNet Lexical Database*

The automatic detection of informative contents in text documents using concepts instead of terms presupposes the existence of machine-readable lexical resources both containing terms and their associated concepts. One such resource is WordNet, which is a freely available lexical database in an easy-to-use text format [Fellbaum 1998]. In version 2.1, it contains about of 150.000 content terms (including about 64.000 multi-word terms), 120.000 concepts, 200.000 associations between terms and concepts (as senses of the terms), and 300.000 conceptual relations between concepts. The semantic network comprises relations like synonymy, antonymy, hyperonymy, hyponymy, meronymy, holonymy, and some others. WordNet has been widely used for linguistic processing of texts, e. g. for summarizing and lexical chaining [Barzilay & Elhadad 1997] as well as indexing [Gonzalo & al. 1998].

### 3.1.2 *The IVal System*

IVal<sup>6</sup> is an experimental system designed to provide access to the WordNet lexical database and to use the vast amount of linguistic and conceptual knowledge coded there for text analysis [Reischer 2007a]. The system in its current state consists of several text processing components:

- sentence boundary detection for the dissection of input text into single sentences;
- morphological parser for deflection and decomposition of simplex and complex terms (derivatives and compounds);
- chart parser based multi-word term recognizer for assembling collocations and idioms;
- simple proper name analyser for uninterpretable expressions not recognized by other analysis modules;
- interactive interface to the WordNet database for the expansion of lexical terms and concepts as well as relations between them (domain modeller).

Additionally, the WordNet lexicon has been enlarged for function words and term frequency data. The basic architecture of the system is task-oriented: the browser provides extended access to the WordNet lexical database; the weaver enables the user to define new terms, concepts, and associations; the analyser reads and analyses plain texts.

### 3.1.3 *Thematic Chains*

Topic or thematic chains are an extension of lexical chains [e. g. Barzilay & Elhadad 1997; Silber & McCoy 2002]. Lexical chaining uses the conceptual interwovenness of lexical entries to find semantic relations between terms (or better their associated concepts) in a text. The result is a chain of coreferent terms which represent those concepts of a text that are not only repeated but semantically linked. The strongest chains with respect to length (number of coreferents) and strength (semantic distance/similarity/proximity between the concepts) are a good indication of what a text is about on the conceptual level. In most cases, only synonym, hyperonym, and hyponym relations are used to form coreference chains. However, as [Varelas & al. 2005] and [Budanitsky & Hirst 2006] pointed out, further conceptual relations like antonymy, meronymy, holonymy, and others should be considered. In that case, we must talk of *thematic or topic chains* because the related concepts are not just more or less semantically equal but thematically related. One example may show this: The

---

<sup>6</sup> *Informativeness Evaluator.*

shortest *semantic* distance between the concepts AUTOMOBILE and STEERING-WHEEL is 10, if we simply count all intervening nodes in WordNet 2.1 to step from one concept to the other; in contrast, if we use more relations like the ones given above, the distance reduces to 3.<sup>7</sup>

The automatic construction of thematic chains is basically quite simple: For every identified noun term in the text, all its possible readings (i.e. associated concepts) are tried to be thematically linked to other concepts in existing and open chains. A concept is linked to a chain if its thematic distance to the last N concepts of the chain is below a certain threshold. Concepts not linkable to any existing chain open new chains; chains to which no concept could be linked for a certain period are closed. A chain grows and gets stronger the more thematic relations it can establish to other text concepts; a chain dies if it contains only concepts which are not further supported by the cotext (then it was a wrong thematic thread due to an inadequate interpretation of a term). For example, if the new concept COMPUTER had to be chained to two possible existing chains <PC, NOTEBOOK> or <KEYBOARD, LCD, DRIVE> then there is a strong tendency to link it to one of the chains. But the possible sense 'an expert at calculation' of 'computer' is certainly not linked to any of these chains because it is simply the wrong reading in that context.

In more detail, the procedure for thematic chaining includes several steps:

- Selection of appropriate terms: (i) possible nouns are identified from the text; if a term is ambiguous between two parts of speech (e. g. 'American'), then the term is used as noun because the concepts are strongly related; (ii) very polysemous and frequent noun terms are excluded from further consideration because they seem semantically not discriminative enough (e. g. 'thing'), i.e. they are somehow related to any concept; (iii) for the same reason, concepts of a term having low semantic information contents are excluded [see section 2.3.1]: they are too unspecific or vague (e. g. 'man' in the sense of ADULT MALE is excluded but not as GAME EQUIPMENT CONSISTING OF AN OBJECT USED IN CERTAIN BOARD GAMES).
- Comparison of concepts to chains: the noun concept to be chained is compared against every existing chain and the concepts already included there. For that purpose, the average thematic proximity between the new concept and the concepts already chained has to be measured. If the distance is below a certain

---

<sup>7</sup> The reader may inspect the conceptual path between the two concepts by entering "? automobile & steering wheel" for semantic and "? automobile @ steering wheel" for thematic distance in the IVal browser.

threshold, then the concept is linked to the chain, i.e. extends it. Thematic proximity is calculated as the extent of weighted feature overlap between two concepts, where the features are either conceptually linked neighbour concepts describing the semantothematic vicinity of the two concepts in question, respectively, or the normalized terms extracted from their synsets and glosses.

- Scoring and sorting of chains: the finally resulting chains have to be scored by their length and strength, i.e. the average proximity between the concepts in the chain and the number of concepts included there. A simple measure is the product of these two values (if proximity is maximal at unity). Certain chains must be excluded from the list: e. g. chains containing only one element obviously sustain no relations to other concepts of the text and can be considered as thematic noise. After scoring and sorting out, the most important or representative concepts of the text can be extracted by simply counting the number of their occurrences in all chains. The justification for this procedure is the fact that only concepts with many thematic relations to other concepts are frequently chained so that they must be central to the topic of the text.

The performance of thematic chaining, both quantitatively with respect to processing time and qualitatively with respect to the results obtained, depends heavily on several parameters used in the chaining process. These parameters include

- the maximum frequency and degree of polysemy a noun term may have (relative to the most frequent and polysemous noun) to consider its concepts in chaining at all;
- the minimum semantic information content a concept must have to be used for chaining at all;
- the maximum distance for thematic proximity of two concepts as well as the scoring of the different conceptual relation types (e. g. a hyperonym concept of X is thematically closer to X than a meronym concept of X);
- the maximum number of steps no concept is linked to a chain until it is closed again.

The optimal parameter setting can only be found by experimentation and evaluation which is work in progress; nevertheless, reasonable initial values can be set intuitively to be later adjusted. However, one result is obvious from the beginning: the qualitative performance depends on the density of the conceptual network modelled. The more conceptual relations are available in the lexicon for a certain domain, the better thematic analysis and assessment of texts can be.

Consequently, for further performance improvement additional knowledge resources have to be integrated (see section 3.2 below).

Using thematic chaining for conceptual representation of texts has several advantages:

- Ambiguous terms can be disambiguated: For example, the term 'apple' has several meanings in English, but in the conceptual neighbourhood of MICROSOFT (within one and the same conceptual chain) it is probably to be interpreted as APPLE COMPANY, because thematic proximity between MICROSOFT and APPLE COMPANY is certainly higher than between MICROSOFT and APPLE FRUIT/TREE.
- Automatic segmentation of text: Every thematic chain represents a certain span of text, where more than one chain may cover one and the same passage (due to different interwoven topics within a section). We can simply count the number of chains starting or ending at a certain text position (e. g. at a certain sentence), so that we have a good indication where (sub)topics start and end.
- Rating of concept importance: Concepts appearing in more than one chain sustain several thematic relations to other concepts and may be deemed as central concepts. The more often a concept appears in a chain the more important and central it is. Indirectly, the frequency of a concept is considered because this increases the possibility to be thematically linked to other concepts.
- Rating of text coherence: The less chains a text needs for representation, the more thematically coherent and dense it seems to be. This may be regarded as a quality criterion for the text, e. g. a text is the more understandable the more coherent it is.

In the next section, we will look at some applications of thematic chaining as one form of text representation for informativeness extraction.

#### *3.1.4 Applications*

Detecting and selecting informative contents from a text affects at least three scenarios: extraction of the most important concepts with respect to the central topic(s) of the text (conceptual indexing), extraction of the most informative or most representative passages of the text (informative summary on the basis of the most important concepts), and retrieval of the most informative sections relative to a set of search terms (passage retrieval; see e. g. [Salton & al. 1993]).

Conceptual indexing as extraction of the most important concepts of the text is required primarily for summarizing here: If we take the N most significant concepts and perform a passage retrieval operation (see below) then we gain the most infor-

mative sections (as defined by the text span of a thematic chain) containing the most important concepts. Furthermore, we can use the concept index to reversely create a term index: in WordNet, every concept is linked to a so-called synset (synonym set) which contains all possible terms expressing the concept. The advantage of this procedure is the inclusion of real topic terms (based on topic concepts) automatically excluding thematic noise terms.

The retrieval of passages being informative about some search terms and their concepts employs the same algorithm as used in thematic chaining: all search terms in their possible readings (concepts) are tried to be included in every thematic chain by conceptual distance. The more of the concepts of the terms can be (virtually) chained with the less thematic distance, the better a specific interpretation (combination of concepts) of the search terms matches with a chain. For example, the search terms 'keyboard' and 'drive' have both several possible meanings of which only the combination COMPUTER-KEYBOARD and PC-STORAGE-DRIVE are thematically or conceptually close to the chain <KEYBOARD, LCD, DRIVE>; neither other term readings (e. g. KEY-HOLDER and DRIVEWAY) nor other chains (e. g. <PC, NOTEBOOK>) are appropriate candidates for adequate results.

In ranking, the passages are primarily scored by their total average distance of all terms' best readings. Additionally, the most informative passages with respect to semantic and thematic information content may be scored for understandability (e. g. semantic specificity or density) and local concentration (how many of the text's most important topic terms are included in the passage). The performance of informativeness detection and selection as described above is still a matter of evaluation. First results indicate an approach worth to be further explored. Nevertheless, further experiments have to be conducted: they concern the question, which parameters of the chaining and retrieval process yield optimal performance, because calibration and coordination of all parameters is of major importance for performance. Another question to be answered is the gain in performance by different improvements like pronoun resolution, proper name recognition, better thematic distance measures, as well as WordNet expansion of terms, concepts, and thematic relations.

### 3.2 Prospects

Of course, there are many questions open. We are just at the beginning of automatic detection and selection of informative contents. The transition from formal to content-related text processing is unavoidable: with the advent of the semantic web we need new and better strategies to exploit the conceptual-thematic information

coded in wordnets and ontologies. Terminologically and conceptionally, we should replace the notion 'relevant' by 'informative' indicating the semantic approach to text processing. With ontologies and wordnets becoming continuously larger and better, the retrieval performance will necessarily increase. At the time being, lexical and ontological resources are still far from being perfect with respect to the density of the general and specific knowledge about topic domains. This also limits the application of the procedures described above to well-modelled subdomains. However, in the near future we will have a complete and up-to-date coverage of all terms and concepts in the world, structured into an ontology or WordNet. If such a complete ontology is available it will not vanish anymore.

One possibility to compensate for the current shortcomings is the automatic enlargement of wordnets and ontologies with respect to additional thematic relations. For example, other knowledge resources like the Cyc ontology may be merged with WordNet and enhance qualitative performance. Quantitative performance can be boosted by a complete precalculation of all thematic relation combinations of noun concepts in WordNet (about  $80.000^2 / 2$  concept-concept relations, with 1 byte per pair need 3 GB of memory). The future generations of multi-core CPUs and large memories will increase performance drastically.<sup>8</sup> As a consequence, informativeness evaluation of whole documents and document collections will be possible on every PC.

## **4 Conclusion**

The extraction of informative content units in documents by semantic text analysis is still in its infancy. The IVal system implements one possible approach to concept-based text representation and passage retrieval by exploiting thematic relations between concepts. One benefit of this approach is the effectiveness of implementation: if the procedure for thematic chaining is implemented, we automatically gain a conceptual indexer, summarizer, and passage retriever which are all based on thematic distance measuring. Further experiments must show whether this approach can also be scaled up to document collections.

---

<sup>8</sup> The performance of syntactic term based search cannot be increased anymore, as Google's immediate presentation of retrieval results in the largest document collection of the world impressively proves. The additional processing power should be better invested in semantic search engines.

## 5 Bibliography

- [Barzilay & Elhadad 1997] Barzilay, R. & Elhadad, M. (1997): Using lexical chains for text summarization. In *Proceedings of the Intelligent Scalable Text Summarization Workshop (ISTS'97)*, ACL. <http://citeseer.ist.psu.edu/article/barzilay97using.html> (8.1.2007)
- [Beaugrande & Dressler 1981] Beaugrande de, R.-A. & Dressler, W. (1981): *Introduction to Text Linguistics*. London & New York: Longman.
- [Budanitsky & Hirst 2006] Budanitsky, A. & Hirst, G. (2006): Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1). pp. 13 – 47.
- [Fellbaum 1998] Fellbaum, C. (1998; Ed.): *WordNet – An Electronic Lexical Database*. Cambridge & London: MIT Press.
- [Gonzalo & al. 1998] Gonzalo, J. & Verdejo, F. & Chugur, I. & Cigarran, J. (1998): Indexing with WordNet synsets can improve Text Retrieval. *Proceedings of the COLING/ACL '98 Workshop on Usage of WordNet for NLP*. <http://citeseer.ist.psu.edu/article/gonzalo98indexing.html> (8.1.2007)
- [Fox 1983] Fox, C. J. (1983): *Information and Misinformation. An Investigation of the Notions of Information, Misinformation, Informing, and Misinforming*. Westport & London: Greenwood Press.
- [Halliday & Hasan 1976] Halliday, M. A. K. & Hasan, R. (1976): *Cohesion in English*. London & New York: Longman.
- [Hoey 1991] Hoey, M. (1991): *Patterns of Lexis in Text*. Oxford: University Press.
- [Kintsch & Vipond 1979] Kintsch, W. & Vipond, D. (1979): Reading Comprehension and Readability in Educational Practice and Psychological Theory. In Nilsson, L.-G. (Eds.): *Perspectives on Memory Research*. Hillsdale: Erlbaum. pp. 329 – 365.
- [Mani & Maybury 1999] Mani, I. & Maybury, M. T. (1999; Eds.): *Advances in Automatic Text Summarization*. Cambridge & London: MIT Press.
- [Mizzaro 1997] Mizzaro, S. (1997): Relevance: The Whole History. *JASIS*, 48(9). pp. 810 – 832.
- [Morris & Hirst 1991] Morris, J. & Hirst, G. (1991): Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text. *Computational Linguistics*, 17(1). pp. 21 – 48.
- [Reischer 2006a] Reischer, J. (2006): *Zeichen Information Kommunikation. Analyse und Synthese des Zeichen- und Informationsbegriffs*. <http://www.opus-bayern.de/uni-regensburg/volltexte/2006/740/pdf/ZeichenInfoKomm.pdf> (accessed 8.1.2007)
- [Reischer 2006b] Reischer, J. (2006): IVal – An Alternative WordNet Browser for Evaluating Semantic Informativeness of Concepts. *Proceedings der KONVENS 2006*, pp. 115 – 120. [http://ling.uni-konstanz.de/pages/conferences/konvens06/konvens\\_files/konvens06-proc.pdf](http://ling.uni-konstanz.de/pages/conferences/konvens06/konvens_files/konvens06-proc.pdf) (accessed 14.10.2006)
- [Reischer 2007a] Reischer, J. (2007): *IVal – Informativeness Evaluator for Retrieval*. <http://lingua-ex-machina.de> (accessed 8.1.2007)



- [Reischer 2007b] Reischer, J. (2007): OntoNet – a WordNet-based ontological-lexical development system. (To appear in the Proceedings of the GLDV-07 Workshop on Lexical-Semantic and Ontological Resources, 13.4. – 14.4.2007, Tübingen)
- [Resnik 1995] Resnik, P. (1995): Using Information Content to Evaluate Semantic Similarity in a Taxonomy. *Proceedings of the IJCAI-95*, Vol. I. pp. 448 – 453.
- [Rosch & Mervis 1975] Rosch, E. & Mervis, C. B. (1975): Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7. pp. 573 – 605.
- [Rosch & al. 1976] Rosch, E. & Mervis, C. B. & Gray, W. D. & Johnson, D. M. & Boyes-Braem, P. (1976): Basic Objects in Natural Categories. *Cognitive Psychology*, 8. pp. 382 – 439.
- [Rosch 1978] Rosch, E. (1978): Principles of Categorization. In Rosch, E. & Lloyd, B. B. (1978; Eds.): *Cognition and Categorization*. Hillsdale: Erlbaum. pp. 27 – 48.
- [Salton & al. 1993] Salton, G. & Allan, J. & Buckley, C. (1993): Approaches to passage retrieval in full text information systems. In *ACM SIGIR conference on R&D in Information Retrieval*. pp. 49 – 58. <http://citeseer.ist.psu.edu/salton93approaches.html> (8.1.2007)
- [Shannon 1948] Shannon, C. E. (1948): A Mathematical Theory of Communication. The Bell System Technical Journal, 27. pp. 379 – 423 & 623 – 656. <http://cm.belllabs.com/cm/ms/what/shannonday/shannon1948.pdf> (accessed 14.10.2006)
- [Silber & McCoy 2002] Silber, H. G. & McCoy, K. F. (2002): Efficiently Computed Lexical Chains as an Intermediate Representation for Automatic Text Summarization. *Computational Linguistics*, 28(4). pp. 487 – 496.
- [Tague-Sutcliffe 1995] Tague-Sutcliffe, J. (1995): *Measuring Information. An Information Services Perspective*. San Diego u a.: Academic Press.
- [Varelas & al. 2005] Varelas, G. & Voutsakis, E. & Raftopoulou, P. (2005): Semantic Similarity Methods in WordNet and their Application to Information Retrieval on the Web. *WIDM'05*. pp. 10 – 16.



# Gathering Information in Online Communities: What is the Key to Success?\*

*Christopher Lueg*

School of Computing  
University of Tasmania  
Hobart (TAS), Australia  
*christopher.lueg@utas.edu.au*

## **Abstract**

When investigating information sharing activities in online travel communities we found that the benefit of querying topically related online communities as part of the search process may go well beyond the provision of up-to-date or “complete” information. In particular, we argue that one of the main benefits of approaching an online community for information is the resulting *interactive* process that helps the user *clarify* his or her information needs. Conceptualizing this process as mediated interaction rather than traditional information retrieval suggests to investigate further the specific characteristics of such processes and to explore ways to enable, sustain and support them.

## **Zusammenfassung**

Bei der Untersuchung von informationsverbreitenden Aktivitäten in virtuellen Gemeinschaften fanden wir Hinweise darauf, dass für Informationssuchende die Vorteile des Befragens von virtuellen Gemeinschaften weit über den Wert der erhaltenen Information in Bezug auf Aktualität und Vollständigkeit hinausgehen können. Wir zeigen in diesem Artikel, dass einer der Vorteile oftmals initiierte wechselseitige Informationsaustauschprozesse sind, die dazu führen können, dass Informationssuchende ihre Informationsbedürfnisse besser verstehen und auch eher in der Lage sind, diese auszudrücken. Diese Prozesse als *vermittelte Anfragen* aufzufassen legt nahe, die spezifischen Eigenheiten dieser Prozesse zu studieren und auch nach Möglichkeiten zu suchen, diese zu initiieren, zu erhalten und zu unterstützen.

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 303-312.

## I Introduction

Computer networks have revolutionized the way information is distributed and accessed and have a significant role in transforming society (e. g., [Castells 2001]). Information and communication technologies (ICT) also enabled virtual or online communities. Pioneers of online community development and research Howard Rheingold and Roxanne Hiltz introduced the term 'online community' to connote the intense feelings of camaraderie, empathy and support that they observed among people in the online spaces they studied [Preece and Maloney-Krichmar 2005].

Over the years online communities have been studied extensively (e. g., [Smith and Kollock 1999]). [Preece 2000] discusses ways to establish and sustain online communities. Most relevant to this paper is the observation that online communities are very effective and powerful information sharing habitats (e. g., [Lueg 2001]). Communities of interest [Carotenuto et al 1999] have shown to be places where inquiries by information seekers are usually very welcome.

In this paper we argue that one of the main benefits of approaching an online community for information is the resulting *interactive* process that may help the user *clarify* his or her information needs. We also suggest that such processes are best conceptualized as mediated interaction rather than as querying information systems. We use online tourism communities as example but aim to abstract from interpretations of theme-specific interactions to more general observations regarding the effectiveness of information gathering in online communities.

We proceed as follows. First we introduce online travel communities and the empirical basis of our considerations. Then we discuss a number of information sharing activities we observed and relate them to concepts including *mediation* and *interactivity*. Finally we discuss our findings and outline future work in this area.

## 2 Online Travel Communities and Data Collection

The role of online communities in enhancing tourism has been largely ignored by research [Schwabe and Prestipino 2005] even though travel and tourism in general are topics that receive a lot of attention in online communities. The latter is of interest to information researchers noticing that information sharing activities in online communities often circumvent and thus affect the way information is distributed in often tightly controlled industries. A number of authors argue that

online tourism communities are changing the way travel information is accessed and shared in the travel industry (e.g., [Schwabe and Prestipino 2005]; [Lueg 2006]).

We collected the data used in this paper from a well-established online community, URL *www.australien-info.de* founded in 1998 (*www.australien-info.de/ueberdlp.html*). The topics are mostly related to travel in Australia as well as immigration to Australia. Discussions are typically held in German but occasionally postings are written in English. The community features a relatively small number of “regulars”. Like in most online communities participation in discussions tends to be irregular. The web site is operated commercially but the online community appears to have remained largely independent.

Nette Unterkunft in Brisbane ?? Neu	Tommy	19-01-07	165	5	21-01-07 01:05
Frage zur Gulf Air Neu	Flückiger René	19-01-07	188	1	19-01-07 23:06
Canning Stock Route Neu	Bernd Rombold	19-01-07	136	0	19-01-07 18:29
Airbus A380 Singapore-Airlines Neu	Andreas Adam	19-01-07	360	3	21-01-07 10:44
Trainingspartner Melbourne gesucht Neu	Jens76	19-01-07	154	0	19-01-07 08:51
Outdoor-Geschäft Neu	P.Mueller	19-01-07	253	2	19-01-07 15:50
Noch jemand am 9.2. von Frankfurt nach Perth unterwegs? Neu	Petra	19-01-07	133	2	20-01-07 18:56
Aborigines treffen Neu	Mary	19-01-07	304	4	21-01-07 04:47
start 22.1.07 Neu	Stefano	18-01-07	192	2	19-01-07 22:20
Schlangenbissgefahr steigt Neu	Dieter Unger	18-01-07	753	8	20-01-07 00:06
Solar Saline am Ngaloo Reef geplant... Neu	Dieter Unger	18-01-07	203	0	18-01-07 09:16
An du Spezialisten Neu	Anne März	17-01-07	590	8	21-01-07 16:35
Erfahrungen als Krankenschwester in AUS Neu	Desirée	17-01-07	325	11	21-01-07 23:35
Anrufe mit Handy Neu	Gabriele	17-01-07	377	7	21-01-07 12:21
Whithaven Beach und Hill-Inlet Lookout Neu	Gabriele	17-01-07	315	9	21-01-07 17:20

The data collection period of 6 weeks in late 2006 (the screenshot above was taken a few months later in January 2007) corresponds to the community's usual expiry rate. “Expire” is a term denoting the process of removing postings after a certain time. Reasons for expiring postings include saving storage space, reducing information overload and last but not least removing information that is likely to be outdated.

### 3 How Does the Community Respond to Inquiries?

In order to find out about how the community responds to inquiries we looked at a 6-week window providing 139 inquiries attracting 769 answers and/or follow-up postings. In total the postings attracted 41,207 views suggesting strong interest in the community and its topics. We did not analyze topic-drift [Osborne 1998] in great detail because it seems to be an essential part of the socializing aspect of such

communities. The number of distinct contributors is unknown because determining would require a level of access to the community that “regular members” including the authors did not possess.

Discussion threads consist of initial postings (typically questions but also 'informative' postings including news about Australia and personal announcements e. g. “Back in Germany”) plus the above mentioned answers or follow-up postings.

In average questions attracted 5.53 replies. Quite a few inquiries did not attract any responses. 24 of the 139 threads in posted in 2006 attracted at least 10 responses; 3 attracted at least 20 responses. One discussion attracted 50 responses. The active/passive (post/read) in the community ratio is interesting as passive participation or “lurking” is often assumed to be less valuable than active participation. [Nonnecke and Preece 2000] and others [Lueg 2000], however, contest this view.

An established research perspective regarding the informational capacity of online communities is to view them as functionally equivalent to “information systems” (e. g., [Schwabe and Prestipino 2005]). According to this view, travelers query online communities in order to retrieve information. Research questions exploring communities from this perspective include but are not limited to up-to-date-ness of information, coverage, completeness, response time, etc. A major question is how information made available by online communities compares to information provided by professional guidebooks. Findings (e. g., [Prestipino et al 2006]) may have significant impacts on commercial travel guidebooks currently published as books.

We are particularly interested in questions regarding the social and also the cognitive processes that enable information seekers to find the information they are looking for.

One dimension of the problem space is the information retrieval insight that users often are unable to rationalize their information needs in such a way that they can articulate them in computer terminology such as Boolean queries. Research in interactive information retrieval (e. g., [Koenemann and Belkin 1996]) and related IR disciplines therefore considers the human information seeker an integral part of the retrieval process and consequently information retrieval systems as “systems for supporting people's interaction with information.” [Belkin 1996, p. 27]. These and similar insights from other disciplines motivated us to explore interactivity and its current or potential role in gathering information from online communities.

The second dimension of the problem space is inspired by ethnographic studies of intermediaries in libraries and other customer support settings. These settings ap-

pear to be far removed from online communities but key insights are highly relevant: the work of intermediaries is often “invisible” and therefore difficult to quantify but nevertheless enormously helpful:

One of the most valuable (and unheralded) services librarians provide is to help clients understand their own needs---a kind of information therapy. Interacting with a reference librarian can be very much like going to a good psychotherapist who, through skillful questioning, gets you to talk about what's really bothering you. [Nardi and O'Day 1999, p. 85].

Summing up we are looking for cues indicating how and to what extent interaction with online communities helps information seekers clarify their information needs and find useful information.

Related work includes [Lueg 2006] describing distinct qualities of community information including mediation, expansion and immediacy. The “mediation” quality denotes that feedback from tourism communities enables users to establish their own, independent perspective on destinations in a way that is not pre-mediated by guidebooks and tourism related media. The “expansion” quality denotes that tourism communities may provide information that is not directly related to their original questions but nevertheless highly relevant. The “immediacy” quality denotes that tourism community may not suffer from the time-space dilemma that guidebook authors usually have to deal with.

Expanding on the information systems view they introduced [Prestipino et al 2007] argue certain features of online communities suggest they may actually be superior information systems: because online communities provide a “natural language”-interface, they argue, there is no need to transform an information need into a formalized query language, thereby losing details of the information need, or using fixed navigational structures, e. g. indices. The community may also act pro-actively and may provide information the asking person did not think of or deem necessary or failed to formulate in his query. [Lueg 2006] shows that communities members often offer information even though they probably know the information is not exactly what the information seeker asked for. One of the examples provided detailed an online community conversation initiated by a user inquiring about a hotel near Perth airport. A community member knowing Perth and the location of Perth airport interpreted the information seeker's situation, concluded that most likely, the tourist is going to stay in Perth for at least a day anyway and recommended to seek a hotel in downtown Perth rather than near the airport. The observation was that in strict IR terms the information provided was not directly relevant to the information seeker's query as he or she specifically inquired about a hotel at the air-

port. The argument was that in such situations, the members' understanding of the information seeker's information need overrode the usual relevance criteria.

### 3.1 Intermediaries

In the previous section we motivated looking for cues indicating how and to what extent interaction with online communities helps users clarify information needs and find information. The metaphor we will use is “intermediary” and argue that the process of identifying the information seeker's state of knowledge by community members engaging in discussions with the information seeker is among the particular strengths of this kind of information gathering. We also argue that the process resembles --at least to some extent-- the interaction between information seekers and (professional) intermediaries in library settings.

In library and information science (LIS) the concept of intermediaries facilitating “mediated interaction” is a powerful information retrieval process that avoids some of the problems of fully automated retrieval systems. Mediated interaction involves an information-seeking user, a skilled human intermediary (typically a librarian) and an information retrieval system. As [Saracevic et al 1997] point out:

Ever since the advent of user modeling by automatic or semiautomatic means in [information retrieval], or for that matter in [artificial intelligence], nothing has come close to matching the extent, complexity, and success of user modeling as done by skillful professional intermediaries in direct contact with the user.

User modeling refers to the construction of a (mental) model of the user's information need by the intermediary. This means the interaction and the growing mutual understanding of the information need by both user and intermediary is central to the process.

### 3.2 Examples of Mediation-Type Processes

In what follows we describe a number of conversations we observed on *australien-info.de*. The conversations illustrate what we mean by “mediated interaction” in the context of online communities.

The first example is a discussion triggered by an inquiry regarding “Hinterland von Cairns: Mareeba, Chillagoe, Innot Hot Springs etc.” (11 answers; initial question posted 3/11/06). The inquiry is about exploring the wider Cairns area (“Hinterland”) and the most interesting places in this area. Later on the same day (3/11/06) a community member suggested visiting the rainforest Tablelands. The next day (4/11/06) the original information seeker responded to the suggestion and asked



specific questions about climate conditions etc. Additional information provided by other community members lead to further questions by the information seeker.

Another example of this type of discussion was triggered by an inquiry about “Route von Alice nach Broome” (posted 24/11/06). The context provided by the information seeker was the intention to travel, in a rental Apollo 4WD [4 wheel drive] Adventure camper van, to Broome from Alice Springs. The specific question was whether the 4WD is suitable for traveling the Gumbarel Hyw (sic) in June or whether it would be better to travel the presumably less interesting Gread Central Road (sic). A community member advised that traveling the route in the 4WD mentioned may be problematic. Moreover he suggested that Apollo's rental conditions may actually exclude traveling this and a number of other 'rough' outback routes. The community member also suggested that the Great Central Route may not be as boring as assumed and suggested to consider traveling the Tananmi Road (another outback track). The inquirer responded that he considered the alternative route and explained why he decided against it and thanked for the Adventure Camper advice.

Other inquiries however did not trigger such interactive processes which leads to the question if there are specific aspects that can be used to characterize successful inquiries. One of the difficulties is defining “success” as receiving feedback from the community depends to some significant extent on the very availability of knowledgeable community members. This means e. g. that regardless of the specific characteristics of an inquiry, community response may be poor during holiday periods in either Germany or Australia (expats answering questions).

What we found is that inquiries to online communities that clearly specify the information sought may be answered in a direct manner and won't trigger an interactive process either.

An example of such an inquiry and the community response is “Bilderladen in ASP” (posted 19/11/06) asking for name and/or address of a specific gallery in Alice Springs. The information sought was provided by the community within 2 days:

### Beitrag am Forum:


Fenster schließen

Bilderladen in ASP

Autor(in): Micha [REDACTED]

Datum: 19-11-06 11:56

Hi,  
ich war vor 6 Wochen in einer Bildergalerie, wo sehr schöne Photographien in Form von grossen Bildern ausgestellt und verkauft wurden. Der Laden liegt in einer kleinen Seitenstrasse der Todd Mall. Wenn man Richtung Norden läuft, ging es mal rechts ab, vor oder nach Red Ochre Grill. Wer kann mir Namen oder Adresse sagen? Vielen Dank an alle Alice- Kenner.  
Cu Micha

 Auf diesen Beitrag antworten


Re: Bilderladen in ASP

Autor(in): Robert [REDACTED]

Datum: 21-11-06 09:19

Hallo Micha,  
Du meinst den Laden von Steve Strike. So weit ich weiss, nennt er sein Unternehmen "Outback Photographics". Du wirst seine Webpage aber sicher über seinen Namen finden. Er bietet dort einige Motive auch Online zum Verkauf an!

Gruß  
Rob

 Auf diesen Beitrag antworten

Re: Re: Bilderladen in ASP

Autor(in): Micha [REDACTED]

Datum: 21-11-06 12:53

Hi Rob  
vielen Dank, liegst voll richtig mit Deinem Hinweis. Man kann sich einfach auf die AUS Spezialisten im Forum verlassen.

[http://www.laenderinfo-forum.de/forum/read\\_window.php?f=3&i=85158&t=85158&soactual=0](http://www.laenderinfo-forum.de/forum/read_window.php?f=3&i=85158&t=85158&soactual=0)

Page 1 of 3

Another example of this type of process is “Bundaberg Rum” (posted 22/11/06). The information seeker inquired about opportunities to buy Bundaberg Rum (an Australian specialty often enjoyed by tourists and sought after once they returned home) in Switzerland and received respective advice within approx. an hour. Another one and a half hours later the inquirer thanked for the information.

Another example is “Digitalkamera im Auto aufladen” (posted 23/11/06) about recharging a digital camera while traveling Australia in a car. Within 13 minutes (sic) the inquiry was answered comprehensively. 16 minutes later the inquirer thanked for the information.

Similarly, “Einfuhr von Schokolade” (posted 22/11/06) asking about bringing German chocolate and Gummibaerchen (a kind of jelly beans) into Australia was answered within 24h.

“Camper für 3” posted 19/11/06 inquiring about a “Hitop Camper” camper van and if it is sufficiently roomy for 3 people was answered within 24h as well.

However, inquiries that appear to be seriously under-specified don't seem to attract a lot of responses either. Inquiries such as “I am about to travel from Sydney to Cairns. What should I consider?” are typical examples of this type of inquiry. A possible explanation is that respective information seekers are perceived as not making an effort to ask more precise questions.

## 4 Discussion and Future Research

The main contribution of this paper is that we have provided qualitative analyzes suggesting that some of the most effective information gathering processes in online communities may be best described as “mediated interaction” [Saracevic et al 1997] or “mediated problem solving” rather than just “querying information systems”. This insight is significant as it suggests a distinct type of support for information seekers. The question is not so much how conversations can be “marked up” for efficient retrieval in the information systems sense. Rather, the question is what kind of graphical and/or textual representations can be used to initiate and sustain mediated interaction in online communities? What kind of reward schemes can be introduced?

Most sites supporting the sharing of travel experiences allow posting textual descriptions of routes; some allow posting of pictures, etc but we are not aware of any work looking into the specific requirements of supporting the very “problem solving” that our research suggests to be a core activity of such sites.

When evaluating the information behavior of online communities it is difficult to derive quantitative results as community participation may vary significantly over time for reasons unrelated to the nature of the original inquiries.

We are proceeding in several directions. We are analyzing further data regarding the characteristics of mediated interaction in online communities. Second, we are reviewing the online communities and knowledge management literature to distill recommendations for incentives schemes applicable to interaction support. Third, we are reviewing existing travel community sites as to what means of support they offer (and why they are offered).

## 5 Acknowledgments

The author wishes to thank Gerhard Schwabe and Amanda Spink for stimulating discussions.

## 6 References

- Belkin, N. (1996). Intelligent Information Retrieval: Whose Intelligence? *Proc. Fifth International Symposium for Information Science*, 1996, pp. 25-31, Universitätsverlag Konstanz.
- Castells, M. (2001). *The Internet galaxy. Reflections on the Internet, Business and Society*. Oxford University Press, Oxford, UK.

- Carotenuto, L., Etienne, W., Fontaine, M., Friedman, J., Newberg, H., Muller, M., Simpson, M., Slusher, J., and Stevenson, K. (1999). CommunitySpace: Toward Flexible Support for Voluntary Knowledge Communities. *Proc. Workshop "Changing Places"*, London, UK.
- Koenemann, J., and Belkin, N. (1996). A Case for Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness. *Proc. Human Factors in Computing Systems*. Vancouver BC, Canada, pp. 205 – 212. ACM Press.
- Lueg, C. (2000). Where is the Action in Virtual Communities of Practice? Workshop "Communication and Cooperation in Knowledge Communities" at the German Computer-Supported Cooperative Work Conference, Munich, Germany.
- Lueg, C. (2001). Information Dissemination in Virtual Communities as Challenge to Real World Companies. *Proc. First IFIP Conference on E-Commerce, E-Business, and E-Government (I3E 2001)* Zurich, Switzerland, pp. 261-270. ISBN 0-7923-7529-7.
- Lueg, C. (2006). Mediation, Expansion and Immediacy: How Online Communities Revolutionize Information Access in the Tourism Sector. *Proc. 14<sup>th</sup> European Conference on Information Systems*, Göteborg, Sweden, paper No 418.
- Nardi, B. A. and O'Day, V. L. (1999). *Information Ecologies*. MIT Press, Cambridge MA, USA.
- Nonnecke, B. and Preece, J. (2000). Lurker Demographics: Counting the Silent. *Proc. Annual ACM SIGCHI Conference on Human Factors in Computing Systems*, The Hague, The Netherlands. 2000. pp. 73-80. ACM Press.
- Osborne, L. (1998). Topic Development in USENET Newsgroups. *Journal of the American Society for Information Science*. Volume 49, Issue 11, pp. 1010-1016. Wiley.
- Preece, J. (2000). *Online Communities: Designing Usability and Supporting Sociability*, John Wiley & Sons, Chichester.
- Preece, J., and Maloney-Krichmar, D. (2005). Online Communities: Design, Theory, and Practice. *Journal of Computer-Mediated Communication*, 10(4).
- Prestipino, M., Aschoff, F. and Schwabe, G. (2006). What's the Use of Guidebooks in the Age of Collaborative Media? Empirical Evaluation of Free and Commercial Travel Information. *Proc. 19<sup>th</sup> Bled Conference*.
- Prestipino, M., Aschoff, F. and Schwabe, G. (2007). How up-to-date are Online Tourism Communities? Empirical Evaluation of Commercial and Information Quality. *Proc. HICSS 2007*.
- Saracevic, T., Spink, A., and Wu, M.-M. (1997). Users and Intermediaries in Information Retrieval: What are They Talking About? *Proc. User Modeling*, pp. 43 – 54.
- Schwabe, G., Prestipino, M. (2005). How Tourism Communities can Change Travel Information Quality. *Proc. 13<sup>th</sup> European Conference on Information Systems*.
- Smith, M.A. and Kollock, P., eds (1999). *Communities in Cyberspace*. Routledge, London.

# Content Management mit MyCoRe

## Dokumenten- und Publikationsserver und Archivanwendungen mit der Open Source Software MyCoRe\*

*Wiebke Oeltjen*

Universität Hamburg  
Regionales Rechenzentrum RRZ – MyCoRe Geschäftsstelle  
Schlüterstraße 70  
20146 Hamburg  
*wiebke.oeltjen@uni-hamburg.de*

### **Zusammenfassung**

Dieser Text gibt in Ergänzung zum Poster einen Überblick über das MyCoRe-Projekt und die Open Source Software MyCoRe, die den Software-Kern für individuell zu entwickelnde Anwendungen im Bereich digitaler Bibliotheken, Content Repositorien und Archiven bildet. Des Weiteren werden die Vorteile von MyCoRe aufgezeigt, wie z. B. die Modularität der Software oder die Anpassbarkeit des Metadatenmodells und der Klassifikationssysteme. Abschließend werden Anwendungen aus verschiedenen Bereichen vorgestellt.

### **Abstract**

This abstract gives an overview over the MyCoRe project and the open source software MyCoRe in addition to the poster. The MyCoRe software is used to develop applications for digital libraries, content repositories and archives. An introduction to main advantages of MyCoRe is given, as there is modularity of the software, flexibility in the metadata model as well as user-defined classification systems. Concluding the range of application is named.

## **I MyCoRe – das Projekt**

MyCoRe ist ein Open Source Software Projekt dessen Anliegen es ist, einen Softwarekern (*My Core*) bereitzustellen, mit dem Dokumenten- und Publikations-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 313-316.

server aufgebaut und Archivanwendungen entwickelt werden können (*My Content Repository*) [MyCoRe]. Die Software wird in Kooperation mit mehreren Universitäten und Universitätsbibliotheken entwickelt. Ursprünglich ging MyCoRe aus dem „Multimedialen Lehr- und Lernserver Essen“ *Miless* hervor, der an der Universität Essen entwickelt wurde [Lützenkirchen 02].

Da die Anforderungen der verschiedenen Universitäten und Bibliotheken an einen Dokumentenserver sehr unterschiedlich sind, wurde MyCoRe so entwickelt, dass es flexibel ist und leicht angepasst werden kann. Die Modularität ermöglicht je nach Bedarf Erweiterungen. Mit der Beispielanwendung *DocPortal* ist darüber hinaus ein sofortiger Einsatz der Software möglich.

MyCoRe bietet eine Reihe von Vorteilen für die Anwendungsentwicklung:

- Konfigurierbare Metadatenmodelle, Standard ist der erweiterte Dublin Core (DC) Metadatensatz,
- beliebige hierarchische Klassifikationssysteme, wie z. B. das Dewey Dezimalklassifikationssystem (DDC),
- mehrsprachige Weboberfläche,
- Import und Export von Objekten als XML-Dokument
- Erfassung beliebiger Dateitypen, auch sehr viele oder große Dateien
- Verwaltung ganzer Verzeichnishierarchien je Dokument, z. B. eines HTML-Baumes
- Integration professioneller Audio-/Video-Streaming-Server
- Suche in Metadaten, Volltexten, extrahierten Daten und XML-Strukturen
- Suche über eigene Abfragesprache mit booleschen Operatoren, nach Phrasen oder mit Platzhaltern
- Verteilte Suche über mehrere Dokumenten- bzw. Publikations-Server
- Benutzer-, Gruppen- und Zugriffsrechteverwaltung
- persistente Identifier (URN)
- Unterstützung verschiedener Protokolle und Schnittstellen, wie z. B. Web Services (SOAP), OAI-PMH 2.0 Data Provider (*Open Archive Initiative*), Suchmaschinen- und Z39.50-Schnittstellen, Google Sitemap Protokoll
- Web Content Management System zur Pflege statischer Webseiten
- Bildbetrachter für große Digitalisate und Bilddateien

MyCoRe steht in der Version 1.3 auf den Projektseiten bei [SourceForge] zum Herunterladen zur Verfügung.

## 2 MyCoRe-Anwendungen

In den letzten zwei Jahren sind eine Reihe von MyCoRe-Anwendungen entwickelt worden, die nun online zur Verfügung stehen. Diese sind überwiegend in Universitäts-Bibliotheken zu finden. Folgende Einsatzbereiche werden exemplarisch genannt:

- Ein **Zeitschriftenportal**, das an der Thüringer Universitäts- und Landesbibliothek entwickelt wurde [Journals@Urmel],
- das **Bibliothekssystem**, das die ursprüngliche Software Miles mit MyCoRe vereint [DuEPublico],
- ein **Online-Verlag** [Hamburg University Press],
- ein **Lexikon**, in dem es um die Bereitstellung von bibliografischen Daten geht [LexM],
- die **Archivsysteme**, die Digitalisate von alten Handschriften [Islamische Handschriften] oder Papyri zu Forschungszwecken bereitstellen.

Weitere Anwendungen sind online verfügbar oder in Entwicklung und werden auf der MyCoRe-Webseite [MyCoRe] bekannt gegeben.

## 3 Literaturverzeichnis

DuEPublico: Dokumenten- und Publikationsserver der Universität Duisburg-Essen.

<http://duepublico.uni-duisburg-essen.de/> [10.04.2007]

Hamburg University Press: Online-Verlag der Staats- und Universitätsbibliothek Hamburg.

<http://hup.rrz.uni-hamburg.de/> [10.04.2007]

Islamische Handschriften: <http://islamicmss.dl.uni-leipzig.de/> [10.04.2007]

Journals@Urmel: Zeitschriftenportal der Thüringer Universitäts- und Landesbibliothek.

<http://zs.thulb.uni-jena.de/> [10.04.2007]

LexM: Lexikon verfolgter Musiker und Musikerinnen der NS-Zeit am Musikwissenschaftlichen Institut der Universität Hamburg.

<http://www.lexm.uni-hamburg.de/> [10.04.2007]

Lützenkirchen, F. (2002): MyCoRe – Ein Open-Source-System zum Aufbau digitaler Bibliotheken. *Datenbank Spektrum*, 2(4), November 2002, 23-27.

MyCoRe: Homepage des MyCoRe-Projektes. <http://www.mycore.de/> [10.4.2007]

Papyri: Das Papyrus-Projekt Halle-Jena-Leipzig an der Universität Leipzig, UBL / URZ.

<http://papyri.dl.uni-leipzig.de/> [10.4.2007]

SourceForge: Das MyCoRe Projekt bei SourceForge, der „Open Source software development web site“ von der OSTG (Open Source Technology Group).

<http://sourceforge.net/projects/mycore/> [10.4.2007]

# POSTER

**Datei**

**ISI2007\_Seite316\_PosterOeltjen.pdf**



## **K<sub>3</sub> – Ein kollaboratives Wissensmanagementsystem zur Beförderung von Kommunikations- und Informationskompetenz\***

*Wolfgang Semar, Joachim Griesbaum, Tao Jiang & Rainer Kuhlen*

Universität Konstanz  
Informationswissenschaft  
78457 Konstanz  
*Wolfgang.Semar@uni-konstanz.de*

### **Zusammenfassung**

K<sub>3</sub><sup>1</sup> ist ein offenes, auf Open Source Software basierendes E-Learning System, das in der universitären Ausbildung eingesetzt wird und kollaboratives Erarbeiten von Wissen durch gruppenbasierte elektronische Kommunikationsforen ermöglicht. Die Lernenden erhalten über ein flexibles Gratifikationssystem Anreize, sowohl auf Gruppen- als auch auf der Individualebene, um positive motivationale und kognitive Effekte zu erzielen und somit die Kommunikations- und Informationskompetenz der Teilnehmer zu verbessern.

### **Überwindung von Defiziten bei der Informations- und Kommunikationskompetenz**

Aufgrund der vom BMBF initiierte Studie „Nutzung elektronischer wissenschaftlicher Informationen in der Hochschulausbildung“ (<http://www.stefi.de>) festgestellten gravierende Mängel beim methodisch kontrollierten Umgang mit den Informationsres-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 317-319.

<sup>1</sup> K<sub>3</sub> wird an der Universität Konstanz am Lehrstuhl Informationswissenschaft (Prof. Kuhlen) entwickelt. Es handelt sich dabei um ein vom BMBF (DLR PT-NMB+F) im Rahmen des Programms „Innovation und Arbeitsplätze in der Informationsgesellschaft des 21. Jahrhunderts“ in Bezug auf die Fachinformation gefördertes Projekt (Projektnummer: 08C5896). Weitere Informationen unter <http://www.k3forum.net>.

sources in der Hochschulausbildung wurde mit K<sub>3</sub> ein System entwickelt, das durch die direkte Integration in die Ausbildungspraxis zur Überwindung dieser Defizite beiträgt.

K<sub>3</sub> nutzt die Potenziale eines netzbasierter Wissensmanagementsystems für das individuelle und gruppenbezogene Lernen, indem Kommunikationsforen dazu genutzt werden, wechselseitigen Austausch und Kooperation zwischen den Teilnehmern eines Kurses zu befördern. Hierzu wurden traditionelle Lernmethoden mit netzbasierten wissensgenerierenden Lernmethoden „angereichert“, ein neues Leistungsbewertungssystem, eine Software, die eine Vielzahl von Technologien zur Unterstützung von Wissenskommunikation und Wissensgenerierung zur Verfügung stellt, sowie ein spezielles Gratifikationssystem entwickelt. Auf Lernmethodenebene wird Kollaboration gezielt in Gruppenlernprozessen umgesetzt, in denen die Lernenden in virtuell organisierten Kleingruppen Lernaufgaben im K<sub>3</sub>-System bearbeiten. Diese Rahmenbedingungen befördern zum einen die Informationskompetenz der Teilnehmer, indem zum Zwecke des kollaborativen Aufbaus lehrbezogener Wissensplattformen mit Information kompetent umgegangen werden muss um somit Wissen erwerben bzw. produzieren zu können. Zum anderen wird die Kommunikationskompetenz der Teilnehmer durch Verfahren der Moderation von elektronischen Kommunikationsforen und der Präsentation der Gruppenergebnisse befördert.

Das angesprochene Gratifikationssystem ist ein sozio-technologisches System mit dessen Hilfe das Leistungsverhalten der Teilnehmer erhöht wird. Dies wird nicht durch eine materielle Belohnung als vielmehr durch einen Mehrwert, der über den primären Nutzen (Scheinerhalt) hinausgeht, realisiert. Es geht um den sekundäre Nutzen, der zum einen eine positive Einstellung darüber etwas geleistet zu haben, und zum anderen die Qualifizierung der Fähig- und Fertigkeiten der Teilnehmer, in den Vordergrund stellen. Dabei sind nicht nur die Weiterentwicklung handwerklicher (umgehen mit der eingesetzten Technologie eines kollaborativen Wissensmanagementsystems), sondern auch kognitive Fähig- und Fertigkeiten (Wissenzuwachs) gemeint. Bestandteile dieses Gratifikationssystems sind Feedbackverfahren die den Teilnehmern Rückmeldungen (deskriptive und evaluierende) zu den erzielten Individual- und Gruppenergebnissen und zum Ablauf der Lernprozesse in textueller Form und in Form von Kennzahlen geben. Mit Hilfe eines darauf aufbauenden Bewertungssystems können die Lernenden ihre eigenen Kompetenzen ermesen und gezielt weiter entwickeln. Die Kennzahlen sind durch die Tatsache, dass sie in Echtzeit bereitgestellt werden auch ein Mittel, um die verringerte Wahrnehmung in virtuellen Räumen zu kompensieren. Zugleich können sie aber auch als Analyseinstrument zur Unterstützung der tutoriellen Betreuung genutzt werden. Mit Hilfe dieses Gratifikationssystems kann ganz individuell auf einzelnen Teilnehmer deren

unterschiedlichen Motivationen, Kompetenzen und Fähigkeiten eingegangen werden kann.



# **PsiaOnline – Fachdatenbank für Arzneimittelwechselwirkungen in der psychiatrischen Pharmakotherapie\***

*Anton Köstlbacher<sup>1</sup>, Christoph Hiemke<sup>2</sup>, Ekkehard Haen<sup>3</sup>  
Gabriel Eckermann<sup>4</sup>, Matthias Dobmeier<sup>5</sup>,  
Rainer Hammwöhner<sup>6</sup>*

<sup>1</sup> Lehrstuhl für Informationswissenschaft, Universität Regensburg, 93040 Regensburg, Germany, *anton.koestlbacher@sprachlit.uni-regensburg.de*

<sup>2</sup> Psychiatrische Klinik der Universität Mainz, Untere Zahlbacher Str. 8, 55101 Mainz, Germany, *christoph.hiemke@mail.uni-mainz.de*

<sup>3</sup> Klinik für Psychiatrie und Psychotherapie der Universität Regensburg am Bezirksklinikum, Regensburg, Germany, *ekkehard.haen@medbo.de*

<sup>4</sup> Bezirkskrankenhaus Kaufbeuren, Fachkrankenhaus für Psychiatrie, Psychotherapie, Psychosomatik und Neurologie, Kaufbeuren, Germany, *eckermann@bkh-kaufbeuren.de*

<sup>5</sup> Psychiatrische Tagesklinik Cham der Klinik und Poliklinik für Psychiatrie der Universität Regensburg, Regensburg, Germany, *matthias.dobmeier@medbo.de*

<sup>6</sup> Lehrstuhl für Informationswissenschaft, Universität Regensburg, 93040 Regensburg, Germany, *rainer.hammwoehner@sprachlit.uni-regensburg.de*

## **I Einführung**

Die pharmakologische Behandlung von Patienten mit psychiatrischen Störungen ist oftmals mit einem einzigen Medikament nicht erfolgreich. Eine Kombination verschiedener Psychopharmaka kann deshalb notwendig sein. Auch Kombinationen mit Nichtpsychopharmaka werden angewandt. So benötigt ein Patient der an Depressionen leidet, gleichzeitig aber Herzbeschwerden hat, auch ein Herzmedikament.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 321-326.

Es gibt unzählige mögliche Arzneimittelkombinationen, von denen der größte Teil als sinnvoll und aus Sicht einer Risikoeinschätzung als unkritisch eingestuft wird.

Welche Kombinationen im Einzelfall für den Patienten sinnvoll und ungefährlich sind, ist jedoch wegen der vielen Kombinationsmöglichkeiten dem behandelnden Arzt oft nicht bekannt. Für den niedergelassenen oder in einem Krankenhaus tätigen Arzt ist es daher wichtig, auf aktuelle und verlässliche Informationsquellen zurückgreifen zu können, die ihm konkrete Handlungsanweisungen geben, welche Wirkstoffkombinationen unter welchen Umständen angewendet werden können.

Als Informationsquellen werden dazu in der Regel medizinische Datenbanken genutzt. Es lassen sich einige medizinische Datenbanken anführen, welche mehr oder weniger vollständig die Fachinformationen<sup>1</sup> zu den in Deutschland zugelassenen Präparaten bereit halten. Exemplarisch seien hier nur der Ifap-Index, die Rote Liste und die Informationsangebote von DIMDI genannt.

Die bestehenden Datenbanken oder Datenbanksysteme geben zwar Hinweise auf mögliche Wechselwirkungen (Interaktionen), jedoch nicht wie diese einzustufen sind und wie man die Therapie überwachen kann, wenn Wirkstoffkombinationen nicht vermeidbar sind. Oftmals kann beispielsweise durch Überwachung des Blutspiegels (TDM – Therapeutic Drug Monitoring) eine sinnvolle Wirkstoffkombination unter Vermeidung von Interaktionen angewandt werden.

Es lässt sich also berechtigterweise die Frage stellen, ob die Informationen, welche in diesen Datenbanken abrufbar sind, als ausreichend eingestuft werden können oder ob für den behandelnden Arzt weitere Informationen zu den bekannten Wechselwirkungen bei der Behandlung des individuellen Patienten relevant sein können.

## **2 Projektbeschreibung**

Diese Überlegungen zugrunde legend, entschloss sich eine Arbeitsgruppe aus Ärzten und Naturwissenschaftlern mit Erfahrung auf dem Gebiet der Interaktionspharmakologie zunächst sämtliche möglichen Interaktionen, seien sie in den med. Fachinformationen, in der einschlägigen Literatur, in eigenen Studien oder Einzel-

---

<sup>1</sup> Der Begriff Fachinformation hat in der Pharmazie bzw. Medizin eine spezielle Bedeutung: Es sind damit die nach EU-Direktive 83/570/EEC gesetzlich festgelegten und meldepflichtigen Informationen zu einem Medikament gemeint, welche mit der Zulassung des Medikaments bei den zuständigen Behörden eingereicht werden müssen. Sie sind auch Basis für den Beipackzettel zu einem Arzneimittel.

fallberichten dokumentiert, zusammenzutragen und in einem einheitlichen Format zu erfassen.

Im Unterschied zu anderen Datenbanken dieser Art werden dem behandelnden Arzt konkrete Hilfestellungen gegeben, wie therapeutisch nützliche Arzneimittelkombinationen sicher angewendet werden können und welche Vorsichtsmaßnahmen getroffen werden können, um unerwünschte Interaktionen zu vermeiden.

Die erste Version der Datenbank wurde 2006 als Einzelplatzanwendung vom Springer-Verlag, Heidelberg auf CD-ROM veröffentlicht und ist bereits vergriffen. Sie bietet zu jeder Wechselwirkung detaillierte Informationen über den möglichen Interaktionseffekt und den wahrscheinlichen Interaktionsmechanismus. Darüber hinaus liefert sie eine Bewertung der Wechselwirkung und gibt Hinweise auf die Vorgehensweise zur Handhabung der Interaktion.

In der Datenbank sind knapp 7000 mögliche Wechselwirkungen zwischen fast 1000 verschiedenen Wirkstoffen erfasst. Dabei fällt auf, dass neuere Wirkstoffe oft besser untersucht und deshalb möglicherweise häufiger in der Datenbank vertreten sind. Umgekehrt ist kein Eintrag in der Datenbank keine Garantie dafür, dass nicht doch eine Wechselwirkung mit einem Wirkstoff auftreten kann.

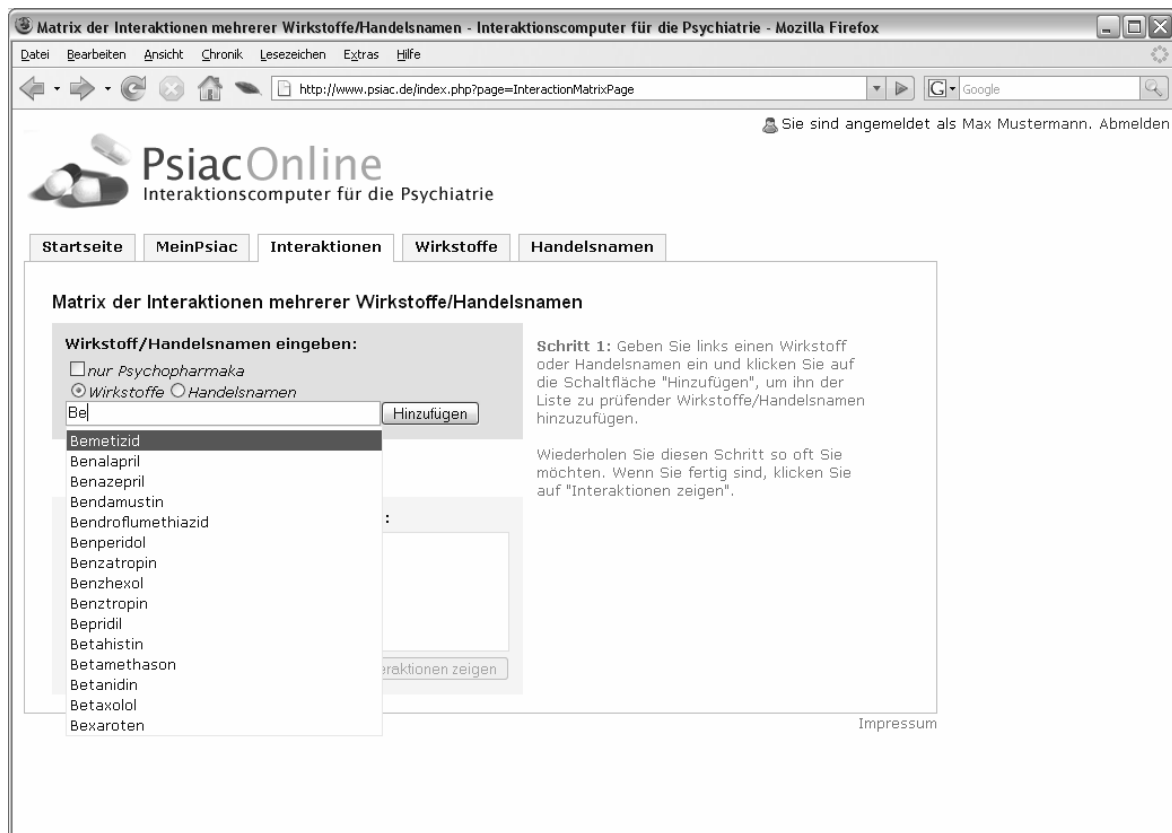
### **3 Aktueller Stand und Ausblick**

Als nächster Schritt folgte Mitte 2006 der Beginn einer transdisziplinären Zusammenarbeit zwischen der Arbeitsgruppe und dem Lehrstuhl für Informationswissenschaft an der Universität Regensburg. Die Implementierung einer Online-Version der Datenbank findet seit diesem Zeitpunkt als Promotionsprojekt unter dem Namen PsiacOnline (<http://www.psiac.de>) an der Universität Regensburg statt.

Derzeit befindet sich das Informationssystem in einer öffentlichen Beta-Phase und ist nach Registrierung kostenlos zugänglich. Das Angebot wird bereits von über 700 Personen aktiv genutzt (Stand: 5. April 2007).

Wesentliche Vorteile der webbasierten Lösung sind die einfache Aktualisierbarkeit des Datenbestandes und die zentrale Wartung der Softwarebasis. Bei der Umsetzung wurde vor allem auf die Rückmeldungen der Benutzer der CD-ROM-Version eingegangen. Konnten bei der CD-ROM nur jeweils zwei Wirkstoffe ausgewählt und deren mögliche Wechselwirkungen überprüft werden, so gibt es bei der Online-Version die Möglichkeit eine Liste von Wirkstoffen und/oder Handelsnamen anzu-

legen, die dann auf alle zwischen ihnen erfassten Interaktionen geprüft werden kann. Auch dies ist eine Funktion, die vergleichbare Systeme so nicht bieten.



Die Personalisierbarkeit ist ein weiterer Zusatznutzen, der sich durch das Angebot als Online-Version bietet. So werden beispielsweise jedem Benutzer seine zehn letzten Suchabfragen nach Wirkstoffkombinationen angezeigt und auch die letzten von ihm betrachteten Wechselwirkungen stehen dem Benutzer auf seiner persönlichen Seite zur Verfügung.

Das Online-Angebot ermöglicht weiterhin die Integration und Erprobung aktueller Webtechnologien, wie sie unter dem Begriff Web 2.0 zusammengefasst werden und damit deren Übertragung auf den Bereich der Fachinformationssysteme.

Dabei ist informationswissenschaftlich von besonderem Interesse, ob sich durch neue Oberflächenkonzepte die Nutzbarkeit von wissenschaftlichen Informationssystemen verbessern lässt und wie weit beispielsweise *Social Software*-Technologien die horizontale und vertikale Kommunikation zwischen den Nutzern und den Autoren eines Informationsangebots verbessern können.

Technologien wie AJAX (Asynchronous Javascript And XML) werden in der Suchfunktion verwendet um das Auffinden von Wirkstoffen und Wechselwirkungen zu



erleichtern, indem nach Eingabe von mindestens zwei Zeichen in das Suchfeld, dem Nutzer ohne erneutes Laden der Seite passende Ergebnisse vorgeschlagen werden.

Im Bereich *Social Software* werden zwei Ansätze verfolgt: Zum einen werden die Nutzer aufgefordert, aktiv Inhalte beizutragen (sog. *user generated content*). Jeder Benutzer des Systems kann öffentliche Kommentare zu den beschriebenen Wechselwirkungen abgeben und hat damit die Möglichkeit die Autoren und die anderen Nutzer auf aktuelle Beobachtungen aus seiner Praxis zeitnah hinzuweisen. Diese Funktion beinhaltet auch die Chance durch Einzelfallberichte auf bisher nicht erfasste Wechselwirkungen hingewiesen zu werden. Dies ermöglicht ggf. unmittelbar die Verbesserung des Datenbestands durch die Autoren der Arbeitsgruppe und bedeutet damit direkten medizinischen Nutzen für die betroffenen Patienten. Zum anderen wird kontinuierlich das Nutzungsverhalten in anonymisierter Form erhoben, um zusätzlich zu den expliziten Beiträgen der Benutzer, auch Informationen darüber zu erhalten, welche Wirkstoffe und Wechselwirkungen im zeitlichen Verlauf für die Nutzer von besonderem Interesse waren.

Erste überblicksartige Betrachtungen der Nutzungsdaten zeigen, dass es zum derzeitigen Stand einige wenige besonders häufig aufgerufene Wechselwirkungen gibt und eine sehr hohe Anzahl an Wechselwirkungen, die überhaupt nur einmal aufgerufen wurden. Hier sind sicherlich die längerfristige Betrachtung und eine größere Zahl an Nutzern notwendig, um mittels geeigneter statistischer Methoden fundierte Erkenntnisse zu gewinnen.

Aktuell werden weitere Funktionalitäten in PsiacOnline implementiert und erprobt. Dazu gehört eine Beobachtungsliste in die jeder Nutzer Wirkstoffe aufnehmen kann (sog. *Bookmarking* von Wirkstoffen), ein Alert-System welches den Nutzer benachrichtigt wenn eine neue Wechselwirkung mit einem der Wirkstoffe in seiner Liste bekannt wird und eine XML-Schnittstelle für den Export aller Inhalte zur Nutzung in anderen Informationssystemen.

## 4 Literatur

- Hiemke, Dobmeier, Eckermann, Haen (2006), Interaktionen in der Psychopharmakotherapie, CD-ROM, Springer Berlin
- Eckermann, Dobmeier, Haen, Hiemke. (2006), IAC, ein Programm zur praktischen Handhabung von Arzneimittelwechselwirkungen in der Psychiatrie in Psychosom. Konsiliarpsychiatr. 1:56-58, Springer Heidelberg

Hammwöhner, Straub (2005), RESIST – Regensburger Signalpfad Informationssystem in  
Proceedings Knowledge eXtended, Schriften des Forschungszentrums Jülich. Reihe  
Bibliothek/Library 14

Deutsches Institut für Medizinische Dokumentation und Information (DIMDI),  
Homepage abgerufen am 5. April 2007 von <http://www.dimdi.de>

ifap Service-Institut für Ärzte und Apotheker GmbH, Homepage  
abgerufen am 5. April 2007 von <http://www.ifap.de>

Rote Liste Service GmbH, Homepage  
abgerufen am 5. April 2007 von <http://www.rote-liste.de>

# Ansatz zu einem konzeptionellen Modell für interaktive Information Retrieval-Systeme mit Unterstützung von Informationsvisualisierung\*

*Paul Landwich<sup>1</sup>, Norbert Fuhr<sup>2</sup> & Matthias Hemmje<sup>1</sup>*

<sup>1</sup> FernUniversität Hagen  
Lehrgebiet Multimedia und  
Internetanwendungen  
58097 Hagen

*{paul.landwich, matthias.hemmje}  
@fernuni-hagen.de*

<sup>2</sup> Universität Duisburg-Essen  
Fakultät 5, IIS  
47048 Duisburg  
*fuhr@uni-duisburg.de*

## **Zusammenfassung**

In diesem Beitrag wird ein neuer Ansatz für die Modellierung und den Entwurf interaktiver Information Retrieval-Systeme vorgestellt, der einerseits eine enge Integration von Syntax, Semantik und Layout der verwalteten Informationsobjekte verfolgt und andererseits den interaktiven Informationsdialog mit Methoden der Informationsvisualisierung unterstützt.

## **1 Einleitung und Motivation**

In der klassischen Information Retrieval-Forschung dominierte in der Vergangenheit die systemorientierte Sicht. Diese Sicht setzt eine explizite Anfrageformulierung voraus und führt damit oft nur zu einem Wiederfinden von schon bekannten und vormals abgespeicherten Daten. Somit trägt sie häufig nicht zur Behebung des der Anfrage zu Grunde liegenden Informationsbedürfnisses bei. Vielmehr sollten Information Retrieval-Systeme neben der reinen Abfrage die kognitiven Fähigkeiten des Benutzers unterstützen, so dass einerseits ein Dialog zwischen Mensch und System realisiert wird und zum anderen die kognitiv günstige Wahrnehmung und

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 327-332.

Steuerung dieses Informationsdialoges, z. B. durch Methoden der Informationsvisualisierung unterstützt wird.

Im Bereich der computergestützten Information Retrieval-Systeme können diese Anforderungen in vielerlei Hinsicht sehr gut durch Methoden der Informationsvisualisierung syntaktischer, semantischer und kognitiver Aspekte von Informationsdialogen und den dazu korrespondierenden dialogorientierten Kontextmengen von Informationsobjekten verschiedener Kategorien realisiert werden.

In diesem Beitrag wird ein grundlegender Ansatz vorgestellt, der als Ausgangspunkt für ein konzeptionelles Modell zum Entwurf von dialogorientierten interaktiven Information Retrieval-Systemen mit Unterstützung von Informationsvisualisierung dient.

## 2 Ansatz eines konzeptionellen Rahmenmodells für ein interaktives Information Retrieval-System

Grundlage des erweiterten Modells (Abb. 1) ist die Existenz von Wissen auf der einen Seite und einem Problem auf der anderen Seite.

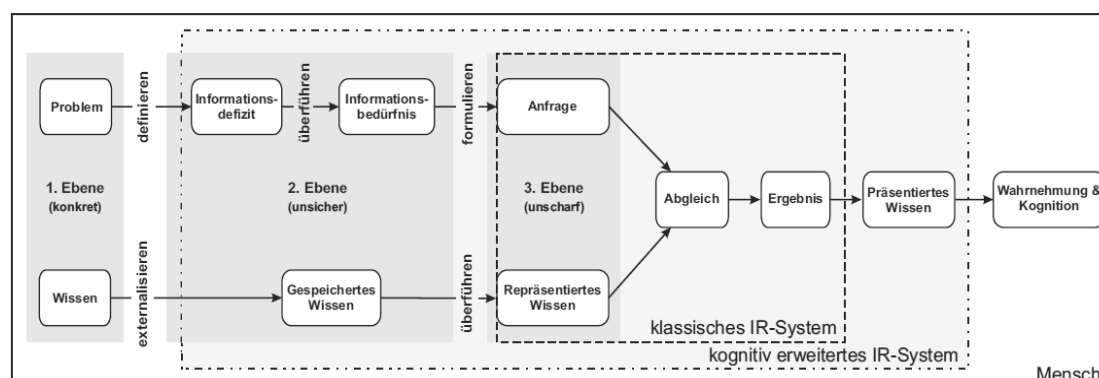
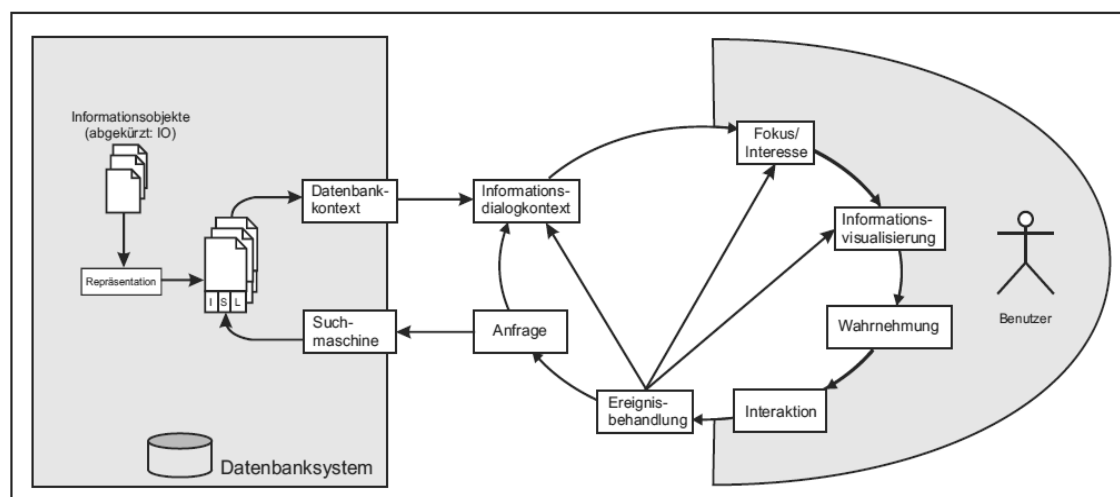


Abbildung 1: Kognitiv erweitertes Modell für Information Retrieval

Der Umgang der Benutzer mit diesen beiden konzeptuellen Ausgangspunkten wird über Prozesse modelliert und diese werden später zusammengeführt. Hierdurch entstehen also zwei nebenläufige Prozessstränge, deren einzelne Stufen aber jeweils drei „Zustände“ (konkret/unsicher/unscharf) durchlaufen. Das erkennbare Ziel für den Entwurf von interaktiven Information Retrieval-Systeme muss es nun sein, die Systemgrenze über die der klassischen Systemfunktionalitäten hinaus zu erweitern, um somit stärker kognitiv orientierte Prozesse der Mensch-Maschine-Interaktion mit einzubeziehen.

Mit einer Integration des konzeptionellen Modells von [Fuhr 92] und des interaktiven Visualisierungsmodells von [Hemmje 99] in ein konzeptionelles Rahmenmodell (s. Abb. 2) entsteht ein neuer Ansatz für die Modellierung sowie den Entwurf und die Spezifikation kognitiv günstiger interaktiver Information Retrieval-Systeme.

Auf der Seite des Datenbank- oder IR-Systems wird zunächst die Erweiterung der Informationsobjekte um die neu definierten Attribute beibehalten. Dafür wird für jedes Informationsobjekt mittels geeigneter Methoden eine entsprechende interne Darstellung erstellt, in der neben Struktur und Layout auch der Inhalt zweckmäßig repräsentiert wird. Eine Anfrage des Benutzers kann somit Selektionsbedingungen enthalten, die sich auf alle drei Aspekte eines Informationsobjekts beziehen. Des Weiteren besteht die Möglichkeit, die Kontextmenge des Informationsdialogs entsprechend geeigneter Attribute innerhalb der dialogorientierten und kognitiven Dimensionen Anfrage,



*Abbildung 2: Rahmenmodell für visuell direkt-manipulatives Information Retrieval*

Ergebnis, Dialog, Historie, Fokus und Kontext zu repräsentieren, zu visualisieren und zum Gegenstand von Interaktion zu machen.

Der verbleibende Teil des Modells auf der Seite der Benutzer zeigt die Einbindung des interaktiven Informationsvisualisierungszyklus. Mit den angebotenen Interaktionswerkzeugen wird z. B. im einfachsten Fall die Möglichkeit der Modifikation der Projektion oder der Ergebnismengendarstellung auf die Dimensionen des Darstellungsraumes (unmittelbare Interaktion) oder z. B. der Reformulierung der Anfrage (mittelbare Interaktion) realisiert. So können z. B. im Laufe eines interaktiven Information Retrieval-Prozesses auf der Grundlage wiederholter Anfrage-Ergebnispräsentationszyklen eine Serie von Teilergebnismengen aus der Gesamtheit aller

Informationsobjekte in der Kollektion durch die Ausführung von Anfragen des Benutzers quasi „exploriert“ und damit in den Dialogkontext aufgenommen werden und stellen somit in ihrer Vereinigung mit den zugrunde liegenden Interaktionen des Dialoges, den korrespondierenden Anfragen und deren strukturellen Beziehungen den Informationsdialogkontext dar.

Der hier skizzierte Ansatz zeigt das große Potential eines solchen Rahmenmodells in groben Zügen auf:

Über die erweiterte Repräsentation der Informationsobjekte und der Möglichkeit, in der Anfrage und in späteren Interaktionen Einfluss darauf zu nehmen, wird eine schärfere Darstellung des vom System gespeicherten Wissens über den Informationsdialog mit den Benutzern ermöglicht.

Der Ansatz über die Integration des Informationsvisualisierungszyklus ermöglicht z. B. eine Unterstützung zur Reduzierung der Unschärfe der Anfragekonstruktion und der Ergebnispräsentation, da über die zyklische Prozessfolge und deren Manipulationsmechanismen eine Kontextmenge des Informationsverhaltens visuell-direktmanipulativ erarbeitet wird, welche das Informationsdefizit verkleinert.

Es ist deutlich sichtbar, dass dem Informationsdialogkontextes eine besondere Bedeutung innerhalb des Modells zukommt. Dieser ist dynamisch und Veränderungen unterworfen, da er immer den aktuellen Explorationszustand eines Dialogs widerspiegelt. Er repräsentiert somit für jeden einzelnen Informationsdialogschritt den Zusammenhang zwischen gespeichertem Wissen und Informationsdefizit und erlaubt die Definition einer Schnittstelle zur benutzerorientierten kognitiv günstigen Kopplung zwischen Datenbank, Suchmaschine und Benutzungsoberfläche.

### **3 Ausblick**

Mit dem vorgestellten ersten Entwurf eines konzeptionellen Rahmenmodells ist die Voraussetzung für die Gestaltung interaktiver Information Retrieval-Systeme mit Unterstützung von Informationsvisualisierung geschaffen. Es gilt nun innerhalb des Rahmenmodells eine Formalisierung und korrespondierende Meta-Modelle für die detailliertere Modellierung der Anfrage-, Informations- und Dialog-, Visualisierungs- und Interaktionsobjekte des Informationskontextes zu erarbeiten, um einen konkreten Informationskontext zwischen Mensch und IR-System zur Laufzeit aufbauen, steuern und verwalten zu können. Beginnend mit exemplarischen Informationsstrategien, wie z. B. anhand von besonders häufig auftretenden Suchstrategien, müssen Klassen von Interaktionsmustern darauf aufbauend repräsentiert und deren

Meta-Aktivitäten auf der taktischen Ebene des Informationsverhaltens mit Hilfe des Rahmenmodells vorgestellt, implementiert und evaluiert werden. Zur Unterstützung dieser dann definierten taktischen Aktivitäten werden danach Informationsvisualisierungen auf der Grundlage einer Analyse der aktuell verfügbaren und häufig verwendeten Visualisierungstechniken erarbeitet.

## **4 Literaturverzeichnis**

- [Fuhr 92] Fuhr, N. Konzepte zur Gestaltung zukünftiger Information Retrieval-Systeme. In Experimentelles und praktisches Information Retrieval: Festschrift für Gerhard Lustig, R. Kuhlen, Ed. Universitätsverlag Konstanz, Konstanz, Germany, 1992, pp. 59 – 75.
- [Hemmje 99] Hemmje, M. Unterstützung von Information-Retrieval-Dialogen mit Informationssystemen durch interaktive Informationsvisualisierung. Dissertation, Darmstadt, 1999.

# POSTER

**Datei**

**ISI2007\_Seite332\_PosterLandwich\_etal.pdf**



# Die Analyse heterogener Unternehmensdatenbestände als Basis für die Visualisierung von Relationen in Suchergebnismengen\*

*Sonja Öttl, Sonja Hierl, Bernard Bekavac & Josef Herget*

SII – Swiss Institute for Information Research  
Hochschule für Technik und Wirtschaft (HTW) Chur  
Ringstrasse/Pulvermühlestrasse 57  
CH-7004 Chur, Schweiz

*{sonja.oettl, sonja.hierl, bernard.bekavac, josef.herget}@fh-htwchur.ch*

## Abstract

Das Projekt „Visual Relations“ verfolgt das Ziel, die Suche in heterogenen Datenbeständen visuell zu unterstützen und Relationen innerhalb der Treffermengen aufzuzeigen. Der Anwender soll hierbei durch geeignete Visualisierungen unterstützt werden, um immanente Strukturen und Verbindungen leichter erkennen zu können. Hierzu soll vor allem die Anzeige von geographischen und zeitlichen Bezügen untersucht werden.

## I Projektkontext

Mit der zunehmenden Digitalisierung von Inhalten aller Arten und der Zunahme von kollaborativen Anwendungen geht ein stetiges Wachstum der jährlich anfallenden Datenmengen in Unternehmen einher, wobei nur ungefähr 20% der gesamten Unternehmensdaten effektiv zur Wertschöpfung genutzt werden [Dragoon 03] und ein durchschnittlicher Arbeitnehmer mit Bürotätigkeiten zwischen 15% und 35% seiner Arbeitszeit bei der Suche nach Informationen verbringt [Feldman 04].

Das Projekt Visual Relations greift diese Problematik auf durch die Entwicklung einer geeigneten Visualisierung zur Darstellung heterogener Datenmengen sowie

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 333-341.

inhärenter Zusammenhänge und Korrelationen von Suchergebnissen unter Berücksichtigung von zeitlichen und geographischen Bezügen.

## **2 Entwicklung von Datenmengen in Unternehmen**

Betrachtet man die Entwicklung bei Unternehmensdaten, die kontinuierlich gespeichert bzw. verwaltet werden, so ist ein stetiges Wachstum zu beobachten. Einer Studie von Forrester Research nach, lagen 2006 weltweit bereits rund 318 Petabyte an Unternehmensdaten vor [Balaouras et al. 2006]. Für die kommenden fünf Jahre wird ein jährliches Wachstum um rund 20%-30% prognostiziert, das bis im Jahr 2011 zu einer Datenmenge von rund 3243 Petabyte führt.

## **3 Herausforderung bei der Integration einer Geografischen Informationssystem-Komponente in die Ergebnisvisualisierung**

Auf dem Markt gibt es bereits einige Suchsysteme, die eine Visualisierung der Ergebnismenge vornehmen, wobei die Dokumentenrepräsentation im Vordergrund steht. In der Regel wird dabei eine Visualisierung der potenziellen thematischen Zusammengehörigkeit von Dokumenten vorgenommen, beispielsweise durch Clustering (zum Beispiel bei Grokker) oder durch den Einsatz von topographischen Metaphern (zum Beispiel bei Kartoo).

Die wenigsten Systeme dieser Art ermöglichen jedoch die visuelle Aufbereitung von zeitlichen oder geographischen Bezügen. Gerade diese Aspekte sind in diversen Anwendungsgebieten, beispielsweise in der Kriminalitätsbekämpfung oder im Bereich der Logistik, jedoch von höchster Relevanz. Ein Grund für fehlende Systemlösungen dieser Art mag in der Komplexität und den Herausforderungen bei der Umsetzung von Zeit- und GIS-Komponenten in Suchsystemen mit Visualisierungskomponenten liegen. In wie weit die Bezüge und Relationen innerhalb von Trefferdokumenten vorhanden sind, wurde zunächst anhand von Testkollektionen untersucht.

## **4 Analyse von Testkollektionen**

Im Rahmen des Projektes wurden vier Testkollektionen untersucht, die bezüglich ihres Homogenitätsgrades sowie ihres Umfangs variieren. Die Untersuchung der

Testkollektionen erfolgte in mehreren Schritten. Nach einer Gesamtsichtung der Dokumente wurden maschinell sämtliche „File System Object“ (FSO)-Attribute ausgelesen und anschließend mittels der Software „R“<sup>1</sup> ausgewertet und visualisiert. Dabei wurden die einzelnen Attribute auf allfällige Korrelationen untersucht, die Zusammensetzung der Kollektionen hinsichtlich Größe, Dateityp etc. fixiert und – sofern sinnvoll – statistische Mittelwerte gebildet.

Im Anschluss wurden die einzelnen Testkollektionen intellektuell erschlossen. Hierzu wurden Stichproben per Zufallsprinzip erhoben, die anschließend gesichtet wurden. Die als zentral erachteten Termini wurden zu jeder Stichprobe ermittelt und in einer Begriffsmatrix notiert. Das Ziel hierbei war, ein Relationsschema zu fixieren, anhand dessen das zu entwickelnde System im weiteren Projektverlauf evaluiert werden kann, da mindestens die intellektuell erschlossenen Relationen auch vom System erkannt werden sollten.

Die Datenbestände der Testkollektionen selbst weisen keine strukturellen Besonderheiten oder signifikante Korrelationen auf. Eine Vielzahl der Dokumente weist nur eine sehr geringe Dateigröße von weniger als 100 KB auf und lediglich einige Ausreißer heben die durchschnittliche Dateigröße stark an.

Versucht man allfällige Relationen zu bestimmen, so müssen zunächst Relationen der Dateien zueinander (z. B. Dateiformat, Erstellungsdatum) und inhaltlichen Relationen (z. B. gleiches Themengebiet) unterschieden werden. Nur bei inhaltlichen Relationen ist die Unterscheidung zwischen strukturierten und unstrukturierten Daten sinnvoll, da maschinelle Auswertbarkeit und somit auch die Qualität der Relationsextraktion bei strukturierten Daten zu wesentlich besseren Ergebnissen führt.

Dateispezifische Attribute können jederzeit automatisiert ausgewertet werden, führen aber nicht zwingend zu zuverlässigen Ergebnissen, da sie durch zahlreiche Prozesse wie das Brennen von Daten auf CDs (Attribut „Date Created“ e. g.) beeinflusst werden können. Der Dateityp sowie das wechselseitige Beinhalten von Dateien (E-Mail-Attachments, ZIP-Dateien, etc.) können als weitestgehend zuverlässig und nachweisbar erachtet werden.

Inhaltliche Attribute sind schwerer zu bestimmen. Das Einbeziehen von Metadaten in die Auswertung der Dateien hat sich als wenig zuverlässig erwiesen, da diese oft nur rudimentär vorhanden oder auch fehlerbehaftet sind. Dementsprechend müssen inhaltliche Verknüpfungen der Dateien primär entweder aus der Dateistruktur oder auch aus den vorkommenden Termini gewonnen werden. Zeitliche und räum-

---

<sup>1</sup> <http://www.r-project.org/>, Stand 20.01.2007.

liche Bezüge können ebenfalls nur in einem geringen Teil der ausgewerteten Dateien eindeutig nachvollzogen werden.

## **5      Potenziell geeignete betriebliche Anwendungen zur Visualisierung geografisch referenzierter Daten**

Die Analyse der Testkollektionen zeigt, zumindest exemplarisch, dass eine Verknüpfung von Sachdaten aus strukturierten bzw. unstrukturierten Unternehmensdatenbeständen mit räumlichen Daten aus Kartenmaterial nicht generell automatisierbar ist bzw. dass die von G10 entwickelte Technologie verfeinert und erweitert werden muss, um angemessene und verwertbare Ergebnisse zu liefern.

Die Auswertung geografischer Beziehungen ist insbesondere für Branchen wie z. B. Transport und Logistik, Reiseveranstaltern oder der Immobilienbranche oder im Rahmen der Fahndung von besonderem Interesse. Um auftretenden Problemstellungen entgegenzuwirken, werden inzwischen vermehrt BI-Software oder auch reine GI-Systeme eingesetzt, die zur Unterstützung des jeweiligen Kerngeschäftes dienlich sind. Diese Systeme benötigen allesamt jedoch stark strukturierte Daten aus Datenbank oder in spezifischen Formaten, um Zusammenhänge hervorheben zu können. Im Bereich der Suchmaschinen und Desktopsuchen dagegen gibt es ad hoc noch keinerlei Produkte, die ähnliche Ziele und Methoden anvisieren.

## **6      Konzept eines Visualisierungssystems**

Die Wahl geeigneter Visualisierungstechniken hängt von der zu Grunde liegenden Datenstruktur ab. Als Ergebnis auf eine Suchanfrage ist zunächst eine netzartige Datenstruktur zu erwarten. Greift man jedoch einen Einzeltreffer heraus und sucht von diesem ausgehend weitere Treffer anhand direkter Relationen, so erhält man eine hierarchische Datenstruktur. Als gängige Visualisierungen für Netze sind zunächst Graphen anzuführen, wobei diese für die Visualisierung umfassender Datenmengen nur eingeschränkt geeignet sind. Hierarchien dagegen werden meist als Baumstruktur visualisiert, beispielsweise in Form eines Hyperbolic Trees [Lamping 1995], eines Cone Trees [Robertson 1991] oder einer Treemap [Johnson & Shneiderman 1991, Shneiderman 1992].

Um diese Visualisierungstechniken geeignet in die Benutzeroberfläche einzubetten, bedarf es einer geschickten Kombination an Interaktionstechniken. Sogenannte „Focus+Context“- oder auch Distortion-Techniques [Leung 1994] wie Fisheye

Views [Furnas 1986] und Bifocal Displays [Spence 1993] e. g. werden oftmals eingesetzt um „Overview and Detail“ [Shneiderman 1996] auf einen Blick zu liefern. Weit verbreitet ist auch der Einsatz von Linking und Brushing, bei dem einzelne Elemente der Visualisierungen miteinander verknüpft interagieren und optisch hierbei hervorgehoben werden. Spätestens seit Raskins Vision einer „Zoomworld“ (vgl. [Raskin 00]) werden auch Zooming-Techniken immer verstärkter angewandt.

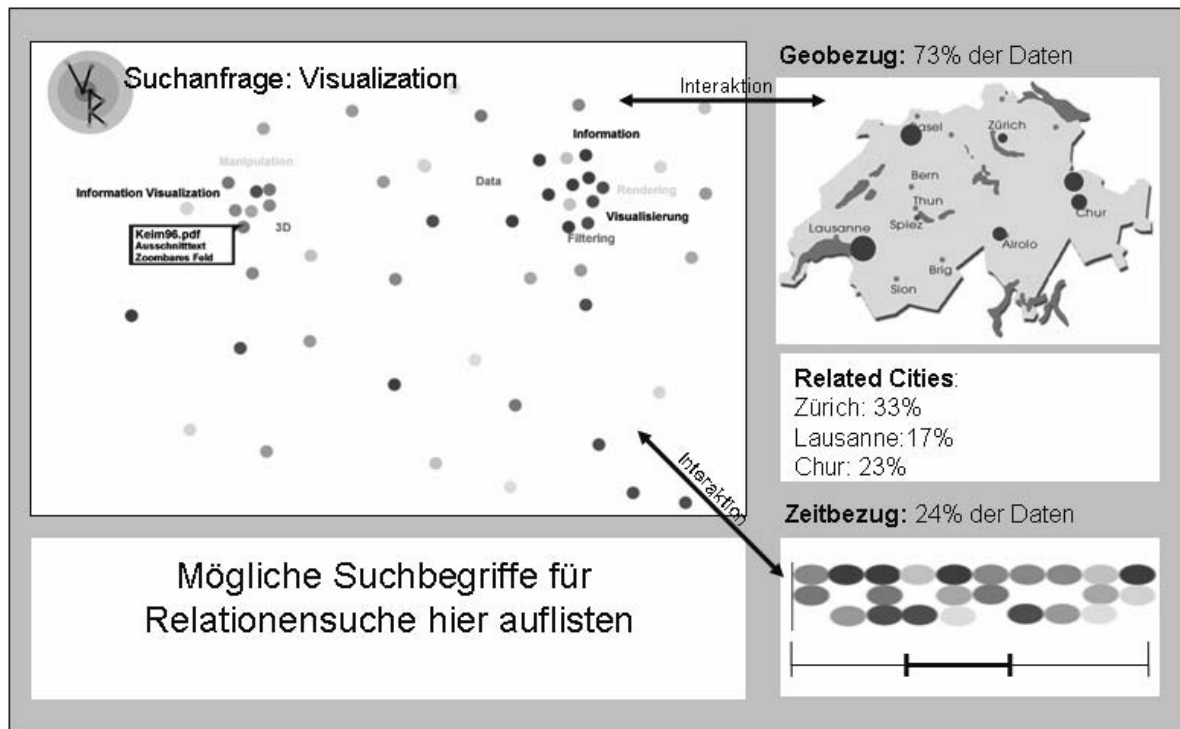


Abbildung 1: Erster Entwurf der Benutzeroberfläche für eine Suchanfrage

Abbildung 1 stellt eine Benutzeroberfläche zur Repräsentation von Suchergebnissen dar. Das Gerüst besteht zunächst aus drei elementaren Modulen: der Document Map (links oben) zur Visualisierung der Treffermenge, dem integrierten GI-System zur Visualisierung geografisch referenzierter Daten (rechts oben) und der Zeitleiste (rechts unten), die unter Verwendung der Fish-Eye-Technik umgesetzt werden soll. Weitere Felder können beispielsweise mit textuellen Ergänzungen oder zusätzlichen Visualisierungen genutzt werden.

Im dargestellten Beispiel ist der jeweilige Dateityp auf die visuelle Variable Farbe gemappt, während die Sättigung die Anzahl der Treffer darstellt. Zudem sind relevante Termini – wie bei den meisten Document Maps üblich – als Landmarken eingeblendet.

Im GI-System repräsentiert die Größe der Kreise die Anzahl der Treffer an einem Ort. Die einzelnen Bereiche interagieren per Linking (Verknüpfung gleicher Elemente in unterschiedlichen Darstellungen) und Brushing (optisches Hervorheben verknüpfter Elemente) miteinander.

Abbildung 2 zeigt die Benutzeroberfläche, die dem Nutzer bei der Suche nach Relationen dienlich sein soll, wobei das Grundgerüst äquivalent zum dargestellten Entwurf in Abbildung 1 aufgebaut wurde. Das Design orientiert sich grundsätzlich an der Metapher einer Zielscheibe, wobei die eingeblendeten Ringe lediglich der Orientierung dienen sollen und in ihrer Anzahl nicht den Grad der Relationen widerspiegeln. Per Mouse-Over-Effekt können Verbindungen zwischen den einzelnen Dateien, die wiederum farblich kodiert wurden, als Graph eingeblendet werden. Da erwartet wird, dass sehr große Datenmengen visualisiert werden müssen, wurde der Einsatz von Fish-Eye-Techniken im Bereich der Relationenvisualisierung als sinnvoll erachtet. Auch hier unterstützen die eingeblendeten Ringe den Nutzer darin, die jeweilige Verzerrung zu erkennen. Ab einem gewissen Grad an Relationen sollen diese nicht mehr direkt präsentiert werden. Pfeile entsprechender Größe oder geeignete Landmarken dienen dem Nutzer an dieser Stelle als Wegweiser, in welche Richtung er sich mittels Panning, dem Verschieben des betrachteten Ausschnitts der Visualisierung, weiterbewegen kann und welche Treffermenge in dieser Richtung zu erwarten ist.

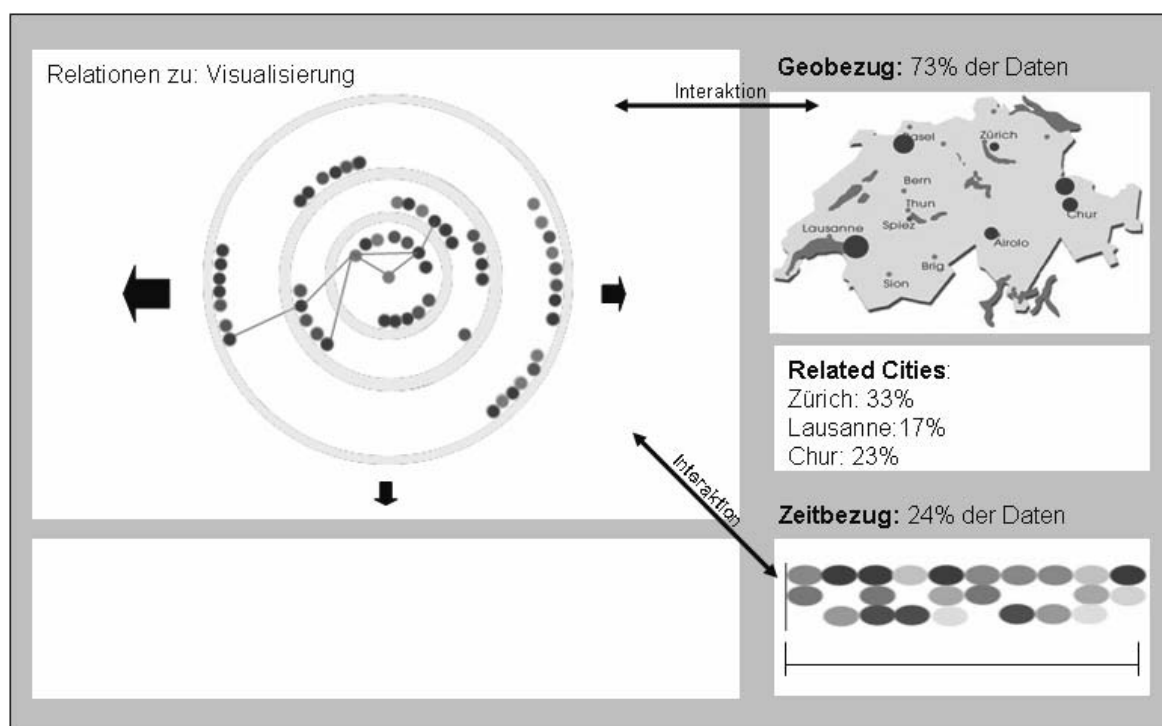


Abbildung 2: Erster Entwurf der Benutzeroberfläche für die Relationensuche

## 7 Ausblick

Unternehmensdatenbestände beinhalten, so das Ergebnis der durchgeführten Analyse, nur zum geringen Teil Dateien, die einen eindeutigen geografischen oder zeitli-

chen Bezug aufweisen. Insbesondere aus unstrukturierten Daten sind diese Attribute schwer zu gewinnen. Entsprechend werden im Rahmen des Projektes Visual Relations verschiedene Visualisierungen miteinander kombiniert, die dem Nutzer unterschiedliche Perspektiven auf die Treffermenge erlauben. Ein weiterer Vorteil des modularen Systemaufbaus liegt beispielsweise darin, Daten mit geografischen Bezug und Daten ohne eindeutigen geografischen Bezug gleichzeitig darstellen zu können und die jeweiligen Visualisierungen interaktiv miteinander zu verbinden. Es verbleibt dem Nutzer selbst im Anschluss an die Recherche, durch „Filtering“-Mechanismen einerseits oder durch „Zooming and Panning“ andererseits, den dargestellten Informationsraum über den Informationszugriff seiner Wahl zu erkunden. Die eingesetzten Visualisierungstechniken müssen ebenso wie die Mechanismen zur Extraktion zeitlicher und örtlicher Relationen – nach einer Evaluation des Prototypen – noch genauer verfeinert werden. Insbesondere durch gezieltes Dato Preprocessing auf Basis einer Auswertung der vom System gefunden Relationen könnte die Qualität der integrierten Visualisierungen wesentlich verbessert werden. Dieser Beitrag führt zu folgenden Erkenntnissen: Geografische Metaphern haben ein Potential, um in der Unternehmenspraxis die Informationsanalyseprozesse zu optimieren, allerdings für eng fokussierte Anwendungsbereiche. Die Analyse der Unternehmensdatenbasis lässt durchaus Attribute extrahieren, die als Grundlage für neue Visualisierungsansätze dienen können. Die vorgestellten Prototypen versprechen einen Lösungsansatz zur skizzierten Problematik und eine verbesserte Interaktion der Nutzer mit Information Retrieval-Systemen zu liefern.

## 8 Literatur

- [Balaouras et al. 2006], Balaouras Stephanie, Schreck Galen, Batiancila Rachel, Disk-Based Data Protection Forecast: 2006 To 2011, Enterprises Shift To Disk As The First Line Of Protection, Forrester Research,  
*<http://www.forrester.com/Research/PDF/0,5110,40036,00.pdf>, 17.11.2006*
- [Dragoon 2003], Dragoon Alice, Business Intelligence Gets Smart(er), Sep. 15, 2003 Issue of CIO Magazine, 2003, <http://www.cio.com/archive/091503/smart.html>, Stand 26.10.06
- [EMC 2006], Homepage der Firma EMC Switzerland, URL:  
*<http://switzerland.emc.com/ilm/>, Stand 12.12.2006*
- [Feldman 2004], Feldman Susan: The high cost of not finding information, March 2004 Issue of KMWorld Magazine, 2004,  
*<http://www.kmworld.com/Articles/ReadArticle.aspx?ArticleID=9534>, Stand 26.10.06*
- [Furnas 1986], Furnas, G. W. 1986. Generalized fisheye views. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Boston, Massachusetts,

- United States, April 13 – 17, 1986). M. Mantei and P. Orbeton, Eds. CHI '86. ACM Press, New York, NY, 16-23.
- [Johnson/ Shneiderman 1991], Johnson, B. and Shneiderman, B. 1991. Tree-Maps: a space-filling approach to the visualization of hierarchical information structures. In Proceedings of the 2nd Conference on Visualization '91 (San Diego, California, October 22 – 25, 1991). G. M. Nielson and L. Rosenblum, Eds. IEEE Visualization. IEEE Computer Society Press, Los Alamitos, CA, 284-291.
- [Keim 2002], Keim, D. A. 2002. Information Visualization and Visual Data Mining. IEEE Transactions on Visualization and Computer Graphics 8, 1 (Jan. 2002), 1-8.
- [Lamping 1995], Lamping, J., Rao, R., and Pirolli, P. 1995. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Denver, Colorado, United States, May 07 – 11, 1995). I. R. Katz, R. Mack, L. Marks, M. B. Rosson, and J. Nielsen, Eds. Conference on Human Factors in Computing Systems. ACM Press/Addison-Wesley Publishing Co., New York, NY, 401-408.
- [Leung 1994], Leung, Y. K. and Apperley, M. D. 1994. A review and taxonomy of distortion-oriented presentation techniques. ACM Trans. Comput.-Hum. Interact. 1, 2 (Jun. 1994), 126-160
- [Lyman/ Varian 2006], Lyman Peter, Varian Hal, How much information (2006), Studie über die weltweite Datenproduktion, <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>, Stand 12.12.06
- [Raskin 2000], Raskin, J. 2000 The Humane Interface: New Directions for Designing Interactive Systems. ACM Press/Addison-Wesley Publishing Co.
- [Robertson 1991], Robertson, G. G., Mackinlay, J. D., and Card, S. K. 1991. Cone Trees: animated 3D visualizations of hierarchical information. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching Through Technology (New Orleans, Louisiana, United States, April 27 – May 02, 1991). S. P. Robertson, G. M. Olson, and J. S. Olson, Eds. CHI '91. ACM Press, New York, NY, 189-194.
- [Schumann/Müller 2000], Schumann H., Müller W., (2000) Visualisierung: Grundlagen und allgemeine Methoden, Berlin
- [Schneiderman 1992], Shneiderman, B. 1992. Tree visualization with tree-maps: 2-d space-filling approach. ACM Trans. Graph. 11, 1 (Jan. 1992), 92-99.
- [Schneiderman 1996], Shneiderman, B. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In Proceedings of the 1996 IEEE Symposium on Visual Languages (September 03 – 06, 1996). VL. IEEE Computer Society, Washington, DC, 336.
- [Spence 1993], Spence, R. 1993. A taxonomy of graphical presentation. In INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems (Amsterdam, The Netherlands, April 24 – 29, 1993). S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel, and T. White, Eds. CHI '93. ACM Press, New York, NY, 113-114



# POSTER

Datei

ISI2007\_Seite353\_PosterOettl\_etal.pdf



# Informationskompetenz als Studienfach?!

## Eine curriculare Zusammenarbeit von Universitätsbibliothek und Fachdisziplinen\*

*Christian Wolff*

Institut für Medien-, Informations- und Kulturwissenschaft  
Universität Regensburg  
93040 Regensburg  
*christian.wolff@sprachlit.uni-regensburg.de*

### **Zusammenfassung**

In diesem Beitrag wird ein seit 2006 an der Universität Regensburg eingeführtes Lehrprogramm für die Schlüsselqualifikation Informationskompetenz vorgestellt. Es zeichnet sich durch die Kooperation von Universitätsbibliothek, Informationswissenschaft und Medieninformatik sowie weiterer Disziplinen aus. Eine weitere Besonderheit ist die Verankerung als Nebenfach im Rahmen der Studiengänge der philosophischen Fakultäten der Universität Regensburg.

### **Abstract**

We discuss a program in information literacy, which has been introduced at Regensburg University in 2006. The program has two main characteristics: The strong cooperation of university library, information and computer science, and humanities who define the support for the information literacy curriculum. At the same time the program is formally established as an integral part of the study programs in the humanities.

## **I Einleitung**

Informationskompetenz dürfte diejenige Schlüsselqualifikation sein, die in fachlicher Perspektive am stärksten aus Bibliotheks- und Informationswissenschaft gespeist wird. Die Auseinandersetzung mit den darunter zu verstehenden Fertigkeiten und Kenntnissen hat vor allem im Bereich US-amerikanischer Bibliotheken bereits

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 343-349.

ein lange Tradition und nicht zufällig wurden einschlägige Standards von der *American Library Association* [ALA 1989; Homann, 2002] formuliert, aber auch im Rahmen des Internationalen Symposiums für Informationswissenschaft ist dieses Thema bereits im Jahr 2000 zum Tagungsmotto erhoben worden [Knorz & Kuhlen, 2000].<sup>1</sup>

In Deutschland wird Informationskompetenz bisher vor allem im bibliothekarischen Umfeld gelehrt; eine Reihe von Universitätsbibliotheken haben Programme und online-Material zu diesem Thema entwickelt (vgl. entsprechende Projekte z. B. an der TU Hamburg-Harburg [Hapke, Bieler, Marahrens, & Rajski, 2003-2005] oder der Zentralbibliothek des Forschungszentrums Jülich [Tunger, Kleikamp, & Salz, (o.J.)]).

An der Universität Regensburg existiert seit dem Sommersemester 2006 ein Studienangebot Informationskompetenz, das sich in mehrerlei Hinsicht von bisherigen Studienmöglichkeiten unterscheidet:

- Seminare zur Schlüsselqualifikation Informationskompetenz können als regulärer Teil des grundständigen Studiums belegt werden, bei gleichzeitiger Integration von bibliothekarischen und fachwissenschaftlichen Angeboten.
- Mit dem Konzept wird eine über den bibliothekarischen Rahmen hinaus erweiterte Sicht auf Informationskompetenz verfolgt, verbunden mit Begleitforschung zur Informationskompetenz.

## 2 Informationskompetenz als Studienfach?!

Die Einrichtung eines Lehrangebotes Informationskompetenz als regulärer Teil eines wissenschaftlichen Studiums (B.A., Magister) ist einer Besonderheit der Studienordnungen der Regensburger philosophischen Fakultäten geschuldet: Der vergleichsweise einfachen Möglichkeit, so genannte „frei kombinierbare Nebenfächer“ definieren zu können, die als Nebenfach in Kombination zu den regulären Studienfächern belegt werden können. Anders als bei den meisten Ausbildungs- und Serviceprogrammen in diesem Bereich ist die Studieneinheit Informationskompetenz vollwertiger Teil des akademischen Studiums.<sup>2</sup>

---

<sup>1</sup> Zur Einführung in die Thematik vgl. [Eisenberg, Lowe, & Spitzer, 2004; Gamble & Easingwood, 2000; Hutchins, Fister, & MacPherson, 2002; Pickering Thomas, 1999].

<sup>2</sup> Das „frei kombinierbare Nebenfach“ bündelt interdisziplinäre und auf Schlüsselqualifikationen ausgerichtete Angebote, neben Informationskompetenz finden sich u. a. Module zu „Sprache und Kommunikation“, „Hypermedia“ oder „Corpuslinguistik“.

Das Aufbaukonzept der Studieneinheit sieht zwei Studienmodule vor, die jeweils im Jahresrhythmus im Sommer- und Wintersemester angeboten werden [Iki, 2006b]. Das Studienprogramm ruht dabei auf drei Säulen:

- Das Seminar- und Übungsangebot der Universitätsbibliothek, das eigens für die Studieneinheit entwickelt wurde, und an dem eine Vielzahl von Fachreferenten der Universitätsbibliothek beteiligt sind.
- Ausgewählte Vorlesungen und Seminare in Informationswissenschaft und Medieninformatik, die einen weiteren Teil des Pflichtprogramms der Studieneinheit ausmachen (u a. die Vorlesungen *Informationsaufbereitung*, *Hypermedia*, *Information Retrieval* und *Multimedia Information Retrieval*).
- Ein semesterweise neu zusammengestelltes Wahlprogramm aus unterschiedlichen, in der Regel geistes- und sozialwissenschaftlichen Fächern (u a. Medienwissenschaft, Medienpädagogik, Medienpsychologie, Philologien, Philosophie, Theologie) das das Kerncurriculum zur Informationskompetenz ergänzt.

Tabelle 1 gibt eine Übersicht zu den im Einführungsseminar und der zugeordneten Übung behandelten Themen.

### 3 Wissenschaftliches Umfeld und Begleitforschung

Die Voraussetzungen für das Studienangebot in Regensburg sind besonders günstig, da an der Regensburger Universitätsbibliothek die Entwicklung innovativer Informationsdienstleistungen eine lange Tradition hat (vgl. z. B. die elektronische Zeitschriftenbibliothek (EZB, [Hutzler, Scheuplein, & Schröder, 2006]) oder das Datenbankinformationssystem DBIS [Hutzler, 2003]). Gleichzeitig ist Regensburg die einzige bayerische Universität, die einen informationswissenschaftlichen Studiengang anbietet, der zuletzt durch das Fach Medieninformatik erweitert wurde. Dabei bringen Informationswissenschaft und Medieninformatik nicht nur Lehrveranstaltungen in die Studieneinheit ein; gleichzeitig können Studenten des Faches Informationswissenschaft durch Belegung der Studieneinheit von dem erweiterten informations- und bibliothekswissenschaftlichen Lehrangebot profitieren.<sup>3</sup>

---

<sup>3</sup> Vereinfachend kann man annehmen, dass so das stärker an Fragen der Recherchepraxis orientierte Curriculum, wie es typischerweise in der Ausbildung von Informationswissenschaftlern an Fachhochschulen vertreten ist, das auf Systementwicklung und Systembewertung ausgerichtete Arbeitsspektrum der Informationswissenschaft an Universitäten ergänzt – eine erfreuliche Koppelung zweier sich fachlich ergänzender Perspektiven.

Grundlagen der bibliothekarischen Informationskompetenz	
Seminarthemen	Übungsthemen
<ul style="list-style-type: none"> <li>• Informationskompetenz und ihre Bedeutung im Wissenschaftsbetrieb (Theorie)</li> <li>• Überblick über Informationsdienstleistungen einer wissenschaftlichen Bibliothek am Beispiel der Universitätsbibliothek Regensburg</li> <li>• Urheberrecht und Bibliothek</li> <li>• Bibliothekarische Sacherschließung</li> <li>• Überblick zu allgemeinen und fachspezifischen Informationsmitteln (elektronisch und konventionell)</li> <li>• Literatursuche in elektronischer Form in überregionalen Katalogen</li> <li>• Recherchestrategien bei der Suche im Internet</li> <li>• Literaturrecherche in Datenbanken</li> <li>• Einführung in Datenbankstrukturen</li> <li>• Nutzung elektronischer Zeitschriften</li> <li>• Literaturverwaltungssysteme im Vergleich</li> <li>• Elektronische Publikationen und elektronisches Publizieren</li> </ul>	<ul style="list-style-type: none"> <li>• Deutsche Verbundsysteme, Sondersammelgebiete und Stellung der Deutschen Nationalbibliothek</li> <li>• Einführung in die Regensburger Verbundklassifikation</li> <li>• Fachübergreifende Zeitschriftendatenbanken</li> <li>• Einbindung von Suchmaschinentechologie in Bibliothekskataloge</li> <li>• Übungen zur Internetrecherche</li> <li>• Recherchestrategien in Datenbanken</li> <li>• Übung zur Nutzung elektronischer Zeitschriften</li> <li>• Literaturverwaltungssysteme</li> <li>• „Open Access“ und die Zeitschriftenkrise</li> <li>• Langzeitarchivierung und Übung zu archivalischen Informationsmitteln</li> </ul>

Tabelle 1: Grundlagen der bibliothekarischen Informationskompetenz

Neben dem Curriculum der Studieneinheit Informationskompetenz entwickelt sich diese Schlüsselqualifikation auch zu einem Forschungsthema in Informationswissenschaft, Medieninformatik, Germanistik und anderen Disziplinen. Seit 2004 wurden mehrere empirische Studien zur Informationskompetenz bei Schülern, Studenten und Rechtsreferendaren durchgeführt [Hochholzer & Wolff, 2005; Mielke & Wolff, 2007], wobei der Schwerpunkt auf der Internetnutzung, Kenntnis einschlägiger Informationssysteme und (mit Einschränkungen) der Recherchekompetenz lag. Eine weitere Studie untersucht Strategien der Informationswissenschaft in forschungsnahen Großunternehmen [Mühlbacher, 2007]. Neben solchen empirischen Studien, die unmittelbar das Thema Informationskompetenz in den Mittelpunkt stellen, sind auch Arbeiten zur Informationsqualität in *social software*-Plattformen wie Wikipedia [Hammwöhner, Fuchs, Kattenbeck, & Sax, 2007] oder zur Entwicklung von Tagging-Systemen im Bereich des *personal information management* [Maßun, 2007] für die Weiterentwicklung der Thematik Informationskompetenz in Lehre und Forschung relevant.

## 4 Offene Probleme und Ausblick

Eine interne Evaluation der Studieneinheit durch die ersten Studierenden im SS 2006 brachte durchweg positive Ergebnisse [Iki, 2006a]. Gleichzeitig belegen empirische Studien zur Informationskompetenz nicht unerhebliche Wissensdefizite im Bereich der Recherchekompetenz (auch bei Dozenten, [Klatt, Gavrilidis, Kleinsimlinghaus, & Feldmann, 2001]) und unterstreichen so die Notwendigkeit einer solchen Studieneinheit. Erstrebenswert ist eine weitere Differenzierung des Angebots, das bisher auf Studiengänge in den Geistes- und Sozialwissenschaften ausgerichtet ist: Die Benutzerstudie zum Portal *vascoda* [Gediga, Gildhorn, & Colver, 2005] zeigt deutlich, dass in verschiedenen Fächergruppen (Medizin, Naturwissenschaften, Geistes- und Sozialwissenschaften) eine sehr unterschiedliche Praxis der Nutzung von Informationsangeboten vorherrscht (Monographien, Zeitschriften, Datenbanken etc.). In einer weitergehenden Perspektive werden Angebote und Methoden des elektronischen Publizierens im Bereich der Wissenschaft (u a. e-Books und e-Prints, Dokumentenserver, Primärdatenveröffentlichung) an Bedeutung und gewinnen und die Anforderung an die Informationskompetenz in der Wissenschaft erhöhen. Erweitert man den Blick über das unmittelbare akademische Umfeld hinaus, so kann man für die durch digitale Informationsangebote geprägte Gesellschaft [Bakardjieva, 2005; Silverstone, 2005] postulieren, dass Informationskompetenz zu einer Schlüsselqualifikation geworden ist, die tatsächlich für jedes Mitglied der Gesellschaft relevant ist [Wolff, 2006].

## 5 Literatur

- [ALA, 1989] American Library Association (ALA). (1989). Presidential Committee on Information Literacy: Final Report. Retrieved April, 2007, from <http://www.ala.org/ala/acrl/acrlpubs/whitepapers/presidential.htm>
- [Bakardjieva, 2005] Bakardjieva, M. (2005). Internet Society. The Internet in Everyday Life. London / Thousand Oaks, CA / New Dehli: Sage Publications.
- [Eisenberg, Lowe & Spitzer, 2004] Eisenberg, M. B., Lowe, C. A., & Spitzer, K. L. (2004). Information Literacy. Essential Skills for the Information Age. New York et al. : Macmillan Publishers.
- [Gamble & Easingwood, 2000] Gamble, N., & Easingwood, N. (Eds.). (2000). ICT and Literacy - Information and Communications Technology, Media, Reading and Writing: London: Continuum.
- [Gediga, Gildhorn & Colver, 2005] Gediga, G., Gildhorn, A., & Colver, B. (2005). Evaluation von vascoda.de aus Benutzersicht. Ergebnisse der Benutzerbefragung 2005.

- Münster / Osnabrück: Institut für wirtschaftspsychologische Forschung und Beratung (IwFB) / Universitäts- und Landesbibliothek Münster.
- [Hammwöhner, Fuchs, Kattenbeck & Sax, 2007] Hammwöhner, R., Fuchs, K.-P., Kattenbeck, M., & Sax, C. (2007). Qualität der Wikipedia – eine vergleichende Studie. In A. Oßwald, M. Stempfhuber & C. Wolff (Eds.), *Open Innovation. Proc. d. 10. Internationalen Symposiums für Informationswissenschaft*, Köln, Mai 2007 (in diesem Band). Konstanz: UVK.
- [Hapke, Bieler, Marahrens & Rajski, 2003-2005] Hapke, T., Bieler, D., Marahrens, O., & Rajski, B. (2003-2005). DISCUS. Online-Tutorial zur Informationskompetenz. Retrieved April, 2007, from <http://discus.tu-harburg.de/login.php>
- [Hochholzer & Wolff, 2005] Hochholzer, R., & Wolff, C. (2005). Informationskompetenz – status quo und Desiderate für die Forschung (e-Print, online verfügbar über den Regensburger Dokumentenserver OPUS, <http://www.opus-bayern.de/uni-regensburg/volltexte/2006/747/>). Regensburg: Universität Regensburg, Institut für Germanistik und Institut für Medien-, Informations- und Kulturwissenschaft.
- [Homann, 2002] Homann, B. (2002). Standards der Informationskompetenz . Eine Übersetzung der amerikanischen Standards der ACRL als argumentative Hilfe zur Realisierung der "Teaching Library". *Bibliotheksdienst*, 36(5), 625 - 637.
- [Hutchins, Fister & MacPherson, 2002] Hutchins, E. O., Fister, B., & MacPherson, K. (2002). Information Literacy Programs, Success and Challenges. In P. Durisin (Ed.), *Changing Landscapes, Enduring Values: Making the Transition from Bibliographic Instruction to Information Literacy* (pp. 3 - 20). Binghamton, NY: The Haworth Press, Inc.
- [Hutzler, 2003] Hutzler, E. (2003). Das Datenbank-Infosystem - eine Dienstleistung kooperierender Bibliotheken. *Bibliotheksforum Bayern (BFB)*, 31(3), 253-260.
- [Hutzler, Scheuplein & Schröder, 2006] Hutzler, E., Scheuplein, M., & Schröder, P. (2006). Der schnelle Weg zum Volltext - Einsatz und Nutzung des Verlinkungsdienstes der Elektronischen Zeitschriftenbibliothek. *Bibliotheksdienst*, 40(3), 306-313.
- [Iki, 2006a] Iki, N. (2006a). Auswertung des Evaluationsbogens zu Seminar und Übung „Grundlagen der Bibliothekarischen Informationskompetenz“ im Rahmen der Studieneinheit INK (Mo1) im SS 2006 durch DozentInnen der Universitätsbibliothek Regensburg (interne Arbeitsmaterialie). Regensburg: Universitätsbibliothek Regensburg.
- [Iki, 2006b] Iki, N. (2006b). Die Regensburger Studieneinheit „Informationskompetenz (information literacy)“ (INK) – eine Kooperation von Bibliothek und Universität. *Bibliotheksdienst*, 40(5), 619-624.
- [Klatt, Gavriilidis, Kleinsimlinghaus & Feldmann, 2001] Klatt, R., Gavriilidis, K., Kleinsimlinghaus, K., & Feldmann, M. (2001). Nutzung elektronischer wissenschaftlicher Information in der Hochschulausbildung. Barrieren und Potenziale der innovativen Mediennutzung im Lernalltag der Hochschulen. Endbericht. (Abschlußbericht Forschungsprojekt Studieren mit elektronischen Fachinformationen (SteFi, <http://www.stefi.de/download/bericht2.pdf>)). Dortmund: Gesellschaft für angewandte Unternehmensforschung und Sozialstatistik (GAUS) mbH / Sozialforschungsstelle Dortmund.



- [Knorz & Kuhlen, 2000] Knorz, G., & Kuhlen, R. (Eds.). (2000). Informationskompetenz – Basiskompetenz in der Informationsgesellschaft. Proc. 7. Intern. Symposium f. Informationswissenschaft, ISI 2000, Darmstadt, November 2000. Konstanz: UVK.
- [Maßun, 2007] Maßun, M. (2007). From Personal Information Management. Towards Collaborative Information Management in Enterprises. In A. Oßwald, M. Stempfhuber & C. Wolff (Eds.), Open Innovation. Proc. d. 10. Internationalen Symposiums für Informationswissenschaft, Köln, Mai 2007 (in diesem Band). Konstanz: UVK.
- [Mielke & Wolff, 2007] Mielke, B., & Wolff, C. (2007 [im Druck]). Juristische Informationskompetenz: Frei verfügbare Web-Quellen versus professionelle Informationsdienstleistungen. In E. Schweighofer & et al. (Eds.), IRIS 2007. Proceedings 10. Internationales Rechtsinformatik Symposium, Salzburg, Februar 2007. Stuttgart et al.: Boorberg.
- [Mühlbacher, 2007] Mühlbacher, S. (2007). Meta-Strategies to foster a Meta-Competence? Supporting information literacy at the workplace In A. Oßwald, M. Stempfhuber & C. Wolff (Eds.), Open Innovation. Proc. d. 10. Internationalen Symposiums für Informationswissenschaft, Köln, Mai 2007 (in diesem Band). Konstanz: UVK.
- [Pickering Thomas, 1999] Pickering Thomas, N. (1999). Information Literacy and Information Skills Instruction, Applying Research to Practice in the School Library Media Center. Westport/CT: Libraries Unlimited.
- [Silverstone, 2005] Silverstone, R. (Ed.). (2005). Media, Technology and Everyday Life in Europe. From Information to Communication. Aldershot / Burlington, VT: Ashgate.
- [Tunger, Kleikamp & Salz, o. J.] Tunger, D., Kleikamp, M., & Salz, E. ((o.J.)). GoInfo. Online-Tutorial der Zentralbibliothek des Forschungszentrums Jülich zum Thema Informationskompetenz. Retrieved April, 2007, from <http://wwwzb.zb.kfa-juelich.de/goinfo/index.htm>
- [Wolff, 2006] Wolff, C. (2006). Information Retrieval is for Everybody - Beobachtungen und Thesen. In M. Schaaf & K.-D. Althoff (Eds.), Lernen, Wissensentdeckung und Adaptivität - LWA / Workshop Information Retrieval der GI - FGIR 2006 (Vol. 1/2006). Hildesheim: Universität Hildesheim.



# **Wissenschaftsportal b2i – Bibliotheks-, Buch- und Informationswissenschaften**

## **Aktueller Stand des Projektes und Ausblick\***

*Elke Bernsee & Christoph Nitrowski*

Fachhochschule Potsdam  
Fachbereich Informationswissenschaften  
Friedrich-Ebert-Straße 4  
14467 Potsdam  
*b2i@fh-potsdam.de, <http://www.b2i.de>*

### **Zusammenfassung**

Das Wissenschaftsportal b2i wird seit Januar 2006 nach dem Vorbild der „Virtuellen Fachbibliotheken“ als DFG-gefördertes Forschungsprojekt auf- und ausgebaut. Im b2i-Portal werden Fachinformation und Datenbanken der drei Wissenschaftsdisziplinen Bibliotheks-, Buch- und Informationswissenschaften gebündelt und über eine gemeinsame Oberfläche zugänglich gemacht. Projektpartner des zunächst für zwei Jahre finanzierten Vorhabens sind: die Staats- und Universitätsbibliothek (SUB) Göttingen als zuständige Sondersammelgebietsbibliothek für das „Informations-, Buch- und Bibliothekswesen“ (SSG 24.1), das Informationszentrum für Informationswissenschaft und -praxis (IZ) der Fachhochschule Potsdam, der „Fachbereich Informationswissenschaften der Fachhochschule Potsdam, das Kompetenznetzwerk für Bibliotheken (KNB) beim Deutschen Bibliotheksverband Berlin und seit dem 1. März 2007 auch die Buchwissenschaft der Friedrich-Alexander-Universität Erlangen-Nürnberg.

### **I      Gegenwärtiger Stand des Projektes**

Ziel dieser ersten Projektphase ist es, relevante Fachressourcen zu bündeln und eine gemeinsame Suchoberfläche aufzubauen. So soll es den Nutzern ermöglicht werden, bequem und komfortabel über eine Metasuchfunktion gleichzeitig in unterschied-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 351-353.

lichen relevanten Fachressourcen zu recherchieren. Interessierte aus Wissenschaft, Studium und Praxis haben so Zugriff auf qualitätskontrollierte Internetquellen, einschlägige Fachdatenbanken, solide Fakteninformation und Kataloge, die gedruckte und elektronische Materialien nachweisen. Damit wird b2i zu einem interdisziplinären „one-stop-shop“ für Fachinformation. Langfristig soll das Wissenschaftsportal auch in Vascoda integriert werden.

Besonderes Anliegen ist es, über das Portal Volltexte anzubieten. In das Portal eingebunden sind bereits jetzt der Potsdamer Volltextserver INFODATA eDepot und der GOEDOC-Dokumentenserver der SUB Göttingen. INFODATA eDepot stellt gegenwärtig bereits über 1500 Dokumente aus dem LIS-Spektrum zur Verfügung. Über ein Eingabeformular können Autoren ihre Publikationen direkt an das IZ übermitteln. Auch über den DINI-zertifizierten GOEDOC-Dokumentenserver der Georg-August-Universität Göttingen (SUB Göttingen) sind LIS-relevante Dokumente abrufbar. Diese sind im Sondersammelgebietskatalog SSG 24,1 der SUB Göttingen verzeichnet.

Von den Bibliografien und Datenbanken, die in die Metasuche integriert werden, sind besonders die bibliografische informationswissenschaftliche Datenbank INFODATA und der Fachausschnitt „Informations-, Buch-, und Bibliothekswesen“ aus der Aufsatzdatenbank Online Contents (OLC) zu nennen. Beide Datenbanken liefern thematisch erschlossene Nachweise auf Artekebene.

Für den Fachbereich Buchwissenschaft entsteht durch die eingebundenen elektronischen Quellen eine erhebliche Verbesserung der Nachweissituation. Bisher in die Metasuche integriert sind: der Katalog des St. Galler Zentrum für das Buch, die Wolfenbütteler Bibliographie zur Geschichte des Buchwesens im deutschen Sprachgebiet (WBB) und die Bibliographische Datenbank zum Bestandserhalt „Buch und Papier“. Die Digitalisierung und Einbindung der Bibliographie zur Buch- und Bibliotheksgeschichte (BBB) Horst Meyers ist derzeit in Arbeit.

Eigens für das Portal aufgebaut wird der „b2i-Guide“, eine Sammlung von intellektuell erschlossenen fachrelevanten Internetquellen, in die der SSG 24,1 Fachausschnitt aus der Elektronischen Zeitschriftenbibliothek (EZB) Regensburg und des Datenbank-Informationssystems (DBIS) eingebunden werden. Auch die Online Kataloge der SUB Göttingen und des IZ/FH Potsdam sind über die Metasuche recherchierbar.

## **2    Ausblick**

Mit diesem Angebot ist das klassische Programm Virtueller Fachbibliotheken in Deutschland abgedeckt. Parallel zur Umsetzung dieser Grundfunktionalitäten wird

unter Einbeziehung der Community an einer erweiterten, nachhaltigen Konzeption für das Wissenschaftsportal b2i gearbeitet. In mehreren Workshops wird gegenwärtig ein Kommunikationsprozess mit der Fachcommunity über Funktionalitäten und mögliche Perspektiven des Portals in Gang gesetzt. So bleibt gewährleistet, dass bei der Weiterentwicklung und beim Ausbau des Portals die Wünsche der Zielgruppen maßgeblich mit einfließen. Die geregelte Einbindung der Interessen der Community und der beteiligten Fachgebiete soll darüber hinaus durch einen wissenschaftlichen Beirat gewährleistet werden.

Ergebnisse eines ersten Workshops während des BID-Kongresses 2007 in Leipzig deuten an, dass die Zielgruppen vor allem im Bereich Studium, Forschung und Lehre gesehen werden. Insbesondere an ein informationswissenschaftliches Portal werden hohe Anforderungen in Bezug auf Innovation und modellorientierte Umsetzung gestellt. Dementsprechend standen auf der Wunschliste ein möglichst umfangreiches Angebot an qualitativ hochwertigen Quellen, verstärkte Serviceleistungen für die Zielgruppen und genormte Schnittstellen zu Web 2.0-Anwendungen. Konkret wurden genannt: personalisierte Dienste, Vernetzung der Datenbankinhalte, der Aufbau einer Experten- und Forschungsdatenbank als Zugang zum Wissen über die Community of Practice, Push-Dienste wie RSS, ein Volltextserver, der auch die Funktion eines fachlichen Repositories erfüllt, ein erweitertes Angebot auch kostenpflichtiger Datenbanken sowie Publikationsmöglichkeiten für Autoren.

Diese Desiderate werden in die Formulierung weiterer Projektziele eingehen. Für die Planungen spielen jedoch nicht zuletzt auch geänderte Rahmenbedingungen eine Rolle, denn zum Jahresende gibt die SUB Göttingen das Sondersammelgebiet 24,1 auf und wird ab 2008 nicht mehr im Projektverbund sein. Eine Neuorientierung von b2i erfolgt dementsprechend auch im Hinblick auf neue Projektpartner und die zukünftige Sondersammelgebietsbibliothek.

### 3 Literatur

- Hobohm, Hans-Christoph; Nitrowski, Christoph: Die Informationswissenschaften in der Virtuellen Fachbibliothek Bibliotheks-, Buch- und Informationswissenschaften "b2i". In: *Content. Proceedings der 28. Online-Tagung / 58. Jahrestagung der DGI*, Frankfurt a.M. 4.-6. Oktober 2006, hrsg. v. M. Ockenfeld, Frankfurt a.M.: DGI, 2006, S. 9-13.
- Weblinks: [www.b2i.de](http://www.b2i.de) (Portal), [b2i.fh-potsdam.de](http://b2i.fh-potsdam.de) (Kommunikationsplattform).



# Semantik in Informationssystemen\*

*Wolfgang Granigg*

Karl-Franzens-Universität Graz,  
Institut für Informationswissenschaft  
und Wirtschaftsinformatik,  
Universitätsstrasse 15/F3,  
8010 Graz, Österreich  
*wolfgang.granigg@gmx.at*

Universität Zürich,  
Institut für Strategie und Unterneh-  
mensökonomik  
Universitätsstrasse 84,  
8006 Zürich, Schweiz  
*wolfgang.granigg@isu.unizh.ch*

Die wissenschaftlichen Disziplinen der Linguistik und der Informationswissenschaft sind von Natur aus auf das engste miteinander verwandt (vgl. [Sparck Jones & Kay 76]). In zahlreichen Gebieten der Informationswissenschaft, wie etwa im zentralen Bereich der Dokumentation, findet man viele linguistische Ansätze wieder und bei einer Betrachtung der jeweiligen Standardlehrbücher dieser beiden Disziplinen fällt auf, dass sehr viele Konzepte einander sehr ähnlich sind bzw. sich stark aufeinander zu beziehen scheinen. Doch nicht nur in die Informationswissenschaft, sondern auch in andere Fachgebiete, die ebenfalls die Betrachtung der Information zum Gegenstand haben, wie etwa die Kognitionswissenschaft oder die Informatik (vgl. [Rauch 94]), haben anscheinend viele Ansätze aus dem Fachgebiet der Linguistik Einzug gehalten.

Insbesondere die Grundlagenwissenschaft der Linguistik, die Semiotik, die für die Auseinandersetzung mit Sprache bzw. mit sprachlichen Aspekten unverzichtbare Begriffe und übergreifende Konzepte liefert (vgl. [Linke, Nussbaumer, Portmann 04]), und konkreter das semiotische Dreieck, das in unterschiedlichsten Interpretationen vorliegt (vgl. [Eco 85]) und unter anderem den Zusammenhang zwischen den drei semiotischen Teildisziplinen, nämlich der Syntaktik, der Semantik und der Pragmatik aufzeigt (vgl. [Maser 71]), stehen dabei im Mittelpunkt unterschiedlicher Theorien und Modelle der verschiedenen oben genannten Disziplinen.

Die abgeschlossene Arbeit setzt sich auf theoretischer Ebene mit dem Themenbereich der ‚Informationssysteme‘ auseinander, und zwar vor allem unter dem Aspekt des „web-basierten Information Retrieval“. Dabei wird im gerade dargestellten in-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 355-359.

terdisziplinären Umfeld versucht, den Themenbereich auch aus unterschiedlichsten Blickwinkeln, aber durch einen gemeinsamen Brennpunkt, nämlich die Semiotik bzw. Semantik, theoretisch zu betrachten.

Ein erstes Ziel der Arbeit ist es mitunter, aufbauend auf semiotischen Überlegungen ein theoretisches Grundmodell eines Informationssystems zu erarbeiten, das modelltheoretische Überlegungen unterschiedlichster Disziplinen verknüpfen kann und im Rahmen von vielen interdisziplinären Forschungsvorhaben als Grund- bzw. Referenzmodell dienen könnte. Aufbauend auf diesem theoretischen Grundmodell eines Informationssystems ist es ein zweites wichtiges Ziel der Arbeit, den Kommunikationsprozess bzw. den Informationsaustausch im Rahmen der Kommunikation eines Menschen bzw. Users mit einem (technischen) Informationssystem schlüssig und auch formalisiert darzustellen. Insbesondere dem begrifflichen Verstehen und damit der Ebene der Semantik sowie der Möglichkeit, aufgrund von Informationen abgeleitete Schlüsse zu ziehen, sollen in diesem Zusammenhang spezielle Aufmerksamkeit geschenkt werden.

Zu Beginn der Arbeit wird ein allgemeiner prägnanter Überblick über grundlegende theoretische Modelle der Daten- und Dokumentenorganisation gegeben. Danach wird von der bis dahin eher konkreten Perspektive ein wenig abgegangen und es werden zentrale Begriffe und Theorieteile wie ‚semantische Relationen‘, ‚Ordnungssysteme‘, ‚Begriffs-‘, ‚Objekt-‘ und ‚Regelräume‘ aufeinander aufbauend eingeführt. Insbesondere der theoretischen Einführung von ‚Ordnungssystemen‘ (vgl. [Gaus 03, Wersig 78, Neveling & Wersig 75]) und einer damit zusammenhängenden informationswissenschaftlichen Fundierung wird dabei größte Aufmerksamkeit geschenkt. Schließlich kann der Begriff eines ‚Informationssystems‘ (vgl. [Wersig 74, Hansen 98]) und somit ein Modell eines solchen auf abstrakter Ebene als ein Zusammenspiel der beiden Elemente ‚bearbeitende Intelligenz‘ und ‚Ordnungssystem‘ festgemacht werden. Darauf aufbauend folgen weitere Ausführungen zum ‚web-basierten Information Retrieval‘. Insbesondere werden hier die idealtypisch zueinander ein wenig konträren Möglichkeiten des web-basierten Information Retrieval, nämlich die Suchabfrage und die Navigation gegenübergestellt (vgl. [Ferber 03, Kuhlen 91, Salton & McGill 87]) und anhand des erarbeiteten Modells eines Informationssystems auf semiotische (bzw. semantische) Stärken und Schwächen hin analysiert.

In weiterer Folge setzt sich die Arbeit auch mit dem ‚Informationssystem Mensch‘ auseinander. Insbesondere wird gezeigt, dass die aus der Kognitiven Psychologie bekannten Ansätze bezüglich der Organisation des Gedächtnisses (vgl. [Anderson 01, Myers 05]) in starke Übereinstimmung zu verschiedenen Überlegungen betreffend die Theorie der Ordnungssysteme gesetzt werden können. Konkret wird



argumentiert, dass auch das ‚Informationssystem Mensch‘ mit dem zuvor erarbeiteten Modell eines Informationssystems beschrieben werden kann. In diesem Zusammenhang werden auch Überlegungen aus dem von Wersig [Wersig 74] beschriebenen informationswissenschaftlichen Modell der Kybiak-Struktur (vgl. [Stachowiak 65, Stachowiak 73]) eingearbeitet.

Schließlich wird die (noch ausständige) Kommunikation zwischen Usern und Informationssystemen unter dem Aspekt des ‚web-basierten Information Retrieval‘ betrachtet. Ein brauchbares Hilfsmodell wird dazu im Kommunikationsmodell von Badura [Badura 71] gesehen, welches auf dem grundlegenden Modell von Shannon und Weaver [Shannon & Weaver 76] aufbaut und dieses in einfacher Form um semiotische Grundelemente erweitert. Dieses Kommunikationsmodell ermöglicht es letztlich, um einige Erweiterungen ergänzt, die verschiedenen erarbeiteten Theorieblöcke in einem erweiterten Modell eines Informationssystems zu vereinen bzw. das erarbeitete Modell eines Informationssystems um Aspekte der Kommunikation zu erweitern. Aufbauend auf diesem erweiterten Modell wird sodann versucht, den Kommunikationsprozess bzw. den Informationsaustausch sowohl verbal, als auch formalisiert schlüssig darzustellen. Im Rahmen dieser Formalisierung des Informationsaustausches wird auf Grundelemente der ‚Theorie der Infone‘ von Devlin zurückgegriffen (vgl. [Devlin 93]), deren Ursprung in früheren Arbeiten zur Situationssemantik von Barwise und Perry zu sehen ist (vgl. [Barwise & Perry 87]).

Zum Abschluss der Arbeit werden schließlich auch noch mögliche ‚dynamische Aspekte in der synthetischen Phase‘ theoretisch betrachtet. Dabei werden – neben informationswissenschaftlichen bzw. vor allem informationsdynamischen Überlegungen (vgl. [Rauch 94a, Rauch 04]) – auch erkenntnistheoretische Überlegungen vorgebracht.

## **Literaturverzeichnis**

- [Anderson 01]. Anderson, J. R. „Kognitive Psychologie“. 3. Auflage. Spektrum Akademischer Verlag, Heidelberg 2001.
- [Badura 71]. Badura, B. „Sprachbarrieren. Zur Soziologie der Kommunikation.“. Friedrich Frommann Verlag, Stuttgart 1971.
- [Barwise & Perry 87]. Barwise, J.; Perry, J. „Situationen und Einstellungen. Grundlagen der Situationssemantik.“. Walter de Gruyter, Berlin 1987.
- [Devlin 93]. Devlin, K. J. „Infos und Infone. Die mathematische Struktur der Information.“. Birkhäuser Verlag, Basel 1993.
- [Eco 85]. Eco, U. „Einführung in die Semiotik“. 5. Auflage. Wilhelm Fink Verlag, München 1985.
- [Ferber 03]. Ferber, R. „Information Retrieval. Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web.“. 1. Auflage. dpunkt Verlag, Heidelberg 2003.

- [Gaus 03]. Gaus, W. „Dokumentations- und Ordnungslehre. Theorie und Praxis des Information Retrieval.“. 4. Auflage. Springer Verlag, Berlin 2003.
- [Hansen 98]. Hansen, H. R. „Wirtschaftsinformatik I. Grundlagen betrieblicher Informationsverarbeitung.“. 7. Auflage. Verlag Lucius und Lucius, Stuttgart 1998.
- [Kuhlen 91]. Kuhlen, R. „Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank.“. Springer Verlag, Berlin 1991.
- [Linke, Nussmauer, Portmann 04]. Linke, A.; Nussbaumer, M.; Portmann, P. R. „Studienbuch Linguistik“. 5. Auflage. Max Niemeyer Verlag, Tübingen 2004.
- [Maser 71]. Maser, S. „Grundlagen der allgemeinen Kommunikationstheorie. Eine Einführung in ihre Grundbegriffe und Methoden (mit Übungen)“. Verlag W. Kohlhammer, Stuttgart 1971.
- [Myers 05]. Myers, D. G. „Psychologie“. Springer Medizin Verlag, Heidelberg 2005.
- [Neveling & Wersig 75]. Neveling, U.; Wersig, G. „Terminologie der Information und Dokumentation“. Herausgegeben vom Komitee Terminologie und Sprachfragen (KTS) der Deutschen Gesellschaft für Dokumentation e.V. (DGD). Verlag Dokumentation, München 1975.
- [Rauch 94]. Rauch, W. „Informationswissenschaft: Theorie für die Praxis.“. In: Der Wirtschaftsingenieur 26 (1994) 1. Österreichischer Verband der Wirtschaftsingenieure, Graz 1994, 8-11.
- [Rauch 94a]. Rauch, W. „Informationsdynamik und Informationspragmatik“. In: Rauch, W.; Strohmeier, F.; Hiller, H.; Schlögl, C. (Hrsg.). „Mehrwert von Information – Professionalisierung der Informationsarbeit“. Proceedings des 4. Internationalen Symposiums für Informationswissenschaft (ISI '94). Universitätsverlag Konstanz, Konstanz 1994, 15-18.
- [Rauch 04]. Rauch, W. „Die Dynamisierung des Informationsbegriffs“. In: Hammwöhner, R.; Rittberger, M.; Semar, W. (Hrsg.). „Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft“. Festschrift für Rainer Kuhlen. UVK Verlagsgesellschaft, Konstanz 2004, 109-117.
- [Salton & McGill 87]. Salton, G.; McGill, M. „Information Retrieval – Grundlegendes für Informationswissenschaftler“. McGraw-Hill, Hamburg 1987.
- [Shannon & Weaver 76]. Shannon, C. E.; Weaver, W. „Mathematische Grundlagen der Informationstheorie“. Oldenbourg Verlag, München 1976.
- [Spark Jones & Kay 76]. Sparck Jones, K.; Kay, M. „Linguistik und Informationswissenschaft“. Verlag Dokumentation, München 1976.
- [Stachowiak 65]. Stachowiak, H. „Denken und Erkennen im kybernetischen Modell“. Springer Verlag, Wien 1965.
- [Stachowiak 73]. Stachowiak, H. „Allgemeine Modelltheorie“. Springer Verlag, Wien 1973.
- [Wersig 74]. Wersig, G. „Information – Kommunikation – Dokumentation. Ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaft“. Beiträge zur Informations- und Dokumentationswissenschaft, Folge 5. 2. Auflage. Verlag Dokumentation, Pullach 1974.
- [Wersig 78]. Wersig, G. „Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis“. DGD-Schriftenreihe, Band 8. Saur Verlag, München 1978.

# POSTER

**Datei**

**ISI2007\_Seite371\_PosterGrannigg.pdf**



# Genome-wide Clustering of Genes\*

*Christian Rengstl*

Universitätsklinikum Regensburg  
Klinik und Poliklinik für Innere Medizin II  
Franz-Josef-Strauß-Allee 11  
93053 Regensburg  
Germany  
*christian.rengstl@klinik.uni-regensburg.de*

## Abstract

With the development of 500K chips, i.e. approximately 500.000 single nucleotide polymorphisms per individual, the quantity and quality of data in genetic researches have risen considerably. As the amount of data makes it hard for any researcher to identify genes and SNPs that are relevant for a specific research problem, it is necessary to organize the data into clusters based on their informativeness. A good approach to cluster the data is to use genetic data, the gene functions and the phenotypes of individuals.

## 1 Introduction

The recent development of so-called 500K chips in the areas of bio-informatics and genetics has enabled researchers world-wide to perform large high-scale studies on the human genome. The idea behind a 500K chip is that for each individual within a population under consideration around 500.000 bi-allelic genetic markers, single nucleotide polymorphisms (SNPs), are genotyped. As SNPs are the largest source of variation in the human genome and have a very low mutation rate in comparison to other genetic markers they are very suited to conduct genetic studies.

These 500K chips, though, produce enormous amounts of data, which makes it hard for the researcher to focus on interesting genes. A lot of applications in the area of bio-informatics calculate the likelihood stating in how far genes might be

---

\* Published in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (eds.) (2007). Open Innovation. Proc. 10<sup>th</sup> International Symposium for Information Science. Constance: UVK, 361-365.

relevant for diseases under consideration. The results can then be organized in signalling pathway networks, i.e. networks of gene interactions, using pathway browsers like Ingenuity Pathway Analysis [<http://www.ingenuity.com>]. Using these networks as starting point, genes and SNPs found on those signalling pathways should be clustered in order to improve the researchers' situation concerning focusing on specific genes among the pool of all input genes.

The input for this project is taken from the KORA (Kooperative Gesundheitsforschung in der Region Augsburg) study which comprises 500K chips of 1644 unrelated individuals. The  $x$  most promising genes will be assembled into signalling pathway networks, which in turn will serve as the input for the clustering procedure.

## 2 Data

The data is stored on a Windows 2003 server running EnterpriseDB version 8.1, which is an extended version of the more known database PostgreSQL. The current size of the database amounts to 143 gigabytes split over around 300 partitioned tables. The database design ensures that all data can be queried through one master table. Each chromosome has in turn an own master table with up to 22 child tables.

SNP data are available in pure text files with the format as seen in table 1. An example for signalling pathway data can be seen in table 2.

<i>Chromosome</i>	<i>Position</i>	<i>Pid</i>	<i>SNP-ID</i>	<i>Allele 1</i>	<i>Allele 2</i>
22	014433758	ZZZ000110011	RS915677	A	C
22	014433758	ZZZ000110022	RS915677	C	T
22	014433758	ZZZ000110033	RS915677	G	G
22	014433758	ZZZ000110044	RS915677	-	-

Table 1: Exemplary SNP data

<i>ID</i>	<i>Genes</i>	<i>Score</i>	<i>Focus Genes</i>	<i>Top Functions</i>
1	ANLN, APAF1, APC, ...	41	35	Cancer, Cellular Assembly and Organization, ...
2	ACP1, ADRB2, ...	41	35	Cellular Compromise, Immune and Lymphatic System Development
3	BCL2, BID, BIRC3, ...	41	35	Cellular Development, Hematological System Development ...

Table 2: Exemplary signalling pathway data

### 3 Planned Approach

In order to cluster this amount of data, it is necessary to minimize the data under consideration before the actual clustering process starts. To accomplish this, it has been planned, and already partially implemented, to calculate the informativeness of all SNPs and genes under consideration. However, before the informativeness of a SNP can be calculated, it is necessary to define the degree of interdependence between SNPs on a gene, the so-called “linkage disequilibrium” (LD). This measure is

calculated using  $r^2 = \frac{D^2}{p_1q_1p_2q_2}$  where

$D = x_{11} - p_1q_1$  with  $x_{11}$  being the frequency of the combination of the first alleles of two SNPs, and  $p_1, p_2, q_1$  and  $q_2$  being haplotype frequencies. LD is in so far important as two SNPs that exceed the threshold of usually 0.8 are considered redundant and therefore can be omitted from the further clustering.

The informativeness of SNPs is defined as the relation of the sum of the entropy of the less frequent to the sum of the entropy of the most frequent allele of all individuals. The entropy of a SNP in turn is defined as  $H_M = -P(A) * \log P(A)$  where  $M$  is a biallelic marker and  $A$  in this case refers to the most frequent allele found on the genetic marker. This way the number of SNPs can be reduced before the clustering process continues. As a signalling pathway can be thought of as a net of links between genes, the PageRank algorithm can be used to deduce the weight of a gene. The informativeness of genes is calculated using an adapted PageRank [Page et al 99] algorithm. In this algorithm all genes are initialized with the sum of the entropy values of all SNPs found on a gene excluding those SNPs that were removed from the initial dataset due to high LD. Like this all genes below a certain threshold, which still has to be defined, can also be excluded from the clustering process.

After the dataset is reduced to only the most relevant items, the actual clustering process can be initiated. As not only the genetic data, i.e. the alleles, of a gene/SNP are interesting for clustering, but also the phenotypical realizations of those genetic data, the phenotypes of the individuals under consideration are also an important part of the clustering process. The aim here is not to cluster genes and SNPs only according to their genetic data but also to include phenotypical data in the clustering. The actual function of a gene within the process of protein synthesis, which can be extracted from protein databases like Swissprot or from the signalling pathway input, will also be considered during clustering. The clustering algorithm that will be implemented for this approach is self-organizing maps. Nevertheless, for evaluation purposes and to combine aspects of several algorithms, more than only one

clustering algorithm will be implemented in order to find the one best suited for the clustering of genetic data in combination with both phenotypes, genetic functions and the informativeness of genes/SNPs.

## 4 References

- [Hampe et al 06] Hampe, Jochen; Schreiber, Stefan; Krawczak, Michael (2006). Entropy-based SNP selection for genetic association studies. *Human Genetics*, 114/1 (2006), 36-43.
- [Page et al 99] Page, Lawrence; Brin, Sergey; Motwani, Rajeev; Winograd, Terry (1999). The PageRank Citation Ranking: Bringing Order to the Web. Stanford Digital Library Technologies Project.  
<http://dbpubs.stanford.edu/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=&name=1999-66.pdf> [January 2007]
- [Zhao et al 05] Zhao, Jinying; Boerwinkle, Eric; Xiong, Momiao (2005). An Entropy-Based Statistic for Genomewide Association Studies. *American Journal of Human Genetics*, 77 (2005), 27-40.
- [Ziegler & König 06] Ziegler, Andreas; König, Inke Regina (2006). A statistical approach to genetic epidemiology: concepts and applications. Weinheim: Wiley-VCH, 2006.



# POSTER

Datei

ISI2007\_Seite377\_PosterRengstl.pdf



# Untersuchung des Online Suchverhaltens im Web

## Interaktion von Internetnutzern mit Suchmaschinen\*

*Nadine Höchstötter*

geb. Schmidt-Mänz, ETU, Universität Karlsruhe (TH), *nsh@topicflux.de*

### **Zusammenfassung**

Es wurden zwei empirische Studien durchgeführt. In der einen Studie konnten Internetnutzer ihr Suchverhalten aus ihrer Sicht in einer Online Umfrage beschreiben. Die Fragen orientieren sich an den Ergebnissen früherer Studien und sind durch neue Fragestellungen bei der Online Suche erweitert worden. In der anderen Studie wurden vier weltweit Datensätze von Suchanfragen generiert. Die Datensätze decken einen Zeitraum von mehr als einem Jahr ab. Erstmals sind hier auch Datensätze von Metasuchern gesondert in Betracht gezogen worden. Auf Basis dieser Daten werden Auswertungsmethoden aus der Literatur aufgegriffen und weiterentwickelt. Fehlerhafte und unvollständige Aussagen in anderen Studien können somit formalisiert und korrigiert werden. Die existierende heterogene Begriffswahl wird zu einer neuen, konsistenten Terminologie zusammengefasst. Anhand von diesen vier unabhängigen Datensätzen werden vollständige Statistiken von Suchanfragen erstellt. Die Formalisierungen können als Grundlage und Vergleichswerte für zukünftige Auswertungen von Suchanfragen herangezogen werden.

### **I Beobachtung von Suchmaschinen-Livetickern**

Es werden bekannte Auswertungsmethoden für Suchanfragen aggregiert und daraus Formalisierungen abgeleitet, welche die Vergleichbarkeit empirischer Datensätze von Suchanfragen gewährleisten. Im Anschluss wird die erhobene Datenbasis vorgestellt und die entsprechenden Kennzahlen berechnet. Die Suchanfragen der vier Suchmaschinen Fireball, Lycos, Metaspinner und Metager sind über einen Zeitraum von mehr als einem Jahr gespeichert worden. Die Suchmaschinen Fire-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informatikwissenschaft. Konstanz: UVK, 367-370.

ball, Lycos und Metaspinner wurden untersucht, da diese einen Liveticker zur Verfügung stellen, in dem die laufend eingehenden Suchanfragen angezeigt werden. Durch ein Programm wurden diese Suchanfragen abgefragt, um sie anschließend zur weiteren Verarbeitung zu speichern. Von Metager wurden die Top-4000 Suchanfragen für diese Erhebung bereitgestellt. Bei der Suchmaschine Lycos, die von den beobachteten Suchmaschinen am stärksten frequentiert ist, umfasst dieser Datensatz rund 190 Millionen Suchanfragen bzw. knapp 340 Millionen Terme.

In der Arbeit wird dargestellt, wie sich das Volumen von Suchtermen im Zeitverlauf mit der Zu- und Abnahme des Interesses der Suchenden für ein bestimmtes Thema verändert. Diese unterschiedlichen charakteristischen Kurven des Volumens von Termen pro Tag werden in grundsätzliche Klassen eingeteilt. Ein Evergreen-Term liegt bspw. in 90 Prozent aller Zeitintervalle über einem definierten Schwellwert (siehe Abbildung 1).

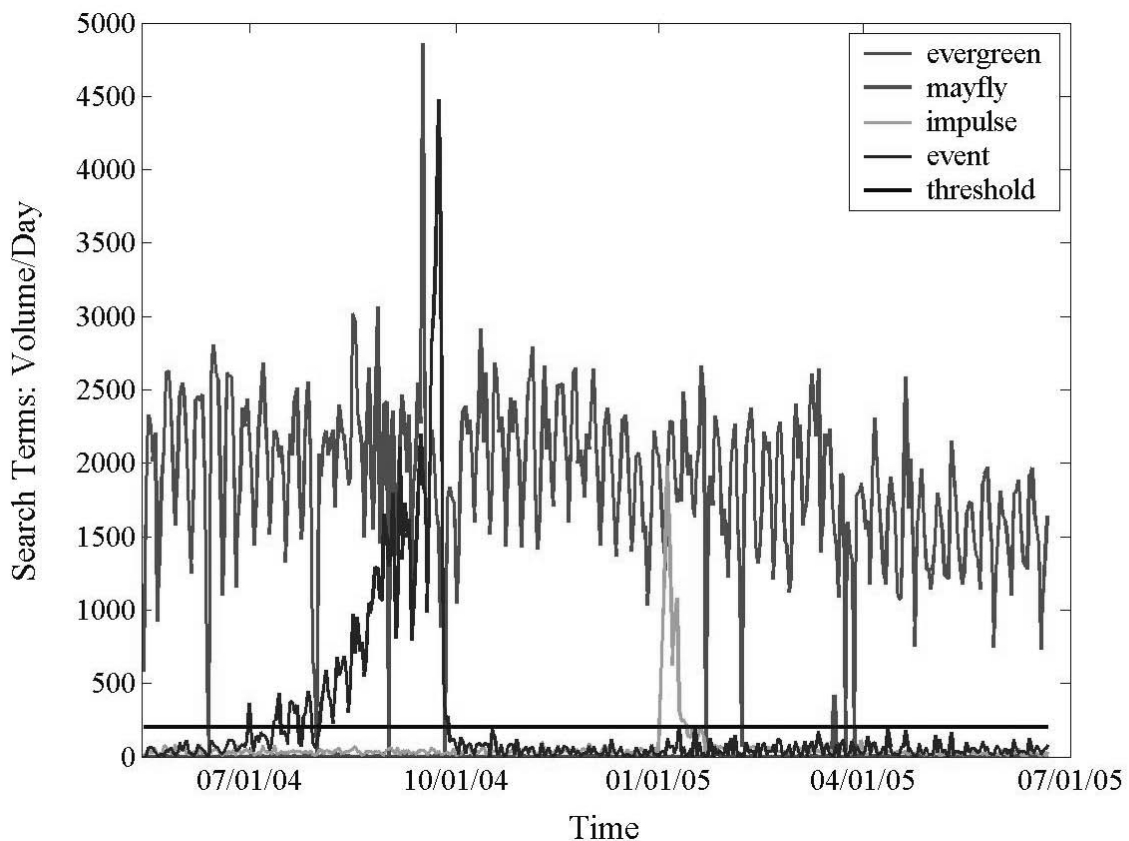


Abbildung 1: Generelle Klassifikation von Suchanfragen

Ein Mayfly-Term bzw. eine Eintagsfliege übersteigt den Schwellwert nur ein oder zwei Mal während der gesamten Erhebungsperiode. Ein Impuls zeigt nach einer Nachricht ein bestimmtes Muster. Direkt nach der Nachricht steigt das Volumen schlagartig an, um nach ein paar Tagen oder einer Woche wieder abzuklingen (Suchbegriffe: Tsunami, Papst). Ein Event ist mit einem wiederkehrenden Ereignis

verbunden und zeigt ein dazu Gegenteiliges Muster. Das Volumen der Suchanfragen steigt sich bis zu dem Tag des Ereignisses, um danach abrupt abzufallen (Suchbegriffe: Weihnachten, Bundestagswahl).

Mit der Kenntnis dieser Verteilungen von Termen ist es möglich, diese auch inhaltlich zu klassifizieren. Bei dem vorgestellten Vorgehen werden die Terme aufgrund der Ähnlichkeit ihrer zeitlichen Struktur, mit der diese Terme nachgefragt werden, inhaltlich klassifiziert. Zeitlich ähnliche Terme können bei der Suche z. B. als Empfehlungen dienen.

## **2 Ausblick**

Insgesamt stellt diese Arbeit durch den Umfang der Literaturstudie, die durchgeführte Umfrage und den Vergleich verschiedener Suchanfragen-Datensätze eine der vollständigsten Arbeiten auf dem Gebiet des Online Suchverhaltens dar. Durch die Formalisierung und Herausarbeitung der Statistiken von Suchanfragen wird eine Einheitlichkeit auf diesem Gebiet geschaffen, die vorher fehlte. Die Analyse von Suchanfragen und vor allem deren zeitabhängige Betrachtung, sind zukunftsweisend für weitere Forschungsansätze in dem Bereich der Trendforschung. Hierbei ist besonders die Betrachtung von Mikrotrends in der Konsumentenverhaltensforschung als ein wichtiger Punkt zu nennen. Die Suchanfragen spiegeln die Bedürfnisstruktur und das steigende und fallende Interesse von Internetnutzern an aktuellen Themen wider. Auch Phänomen wie Mund-zu-Mund-Propaganda werden bei Suchanfragen deutlich. Die weitere Erforschung von Veränderungen in den Volumina von Suchanfragen wird unter dem Begriff „topic flux“ zusammengefasst. Der Begriff „topic“ steht dabei für Thema oder Gesprächsgegenstand, „flux“ bezeichnet die Durchflussmenge (siehe *www.topicflux.de*).

## **3 Literatur**

- Höchstötter, N. (2007): Suchverhalten im Web – Erhebung, Analyse und Möglichkeiten, wird erscheinen in: *Information: Wirtschaft & Praxis*, 58(2007)3.
- Lewandowski, D./ Höchstötter, N. (2007): Qualitätsmessung bei Suchmaschinen, wird erscheinen in: *Informatik Spektrum*, Springer, Heidelberg.
- Lewandowski, D./ Höchstötter, N. (2007): Web Searching: A Quality Measurement Perspective, to be published in: Spink, A./ Zimmer, M. (Eds): *Web Search: Interdisciplinary Perspectives*, Springer, Heidelberg.
- Schmidt-Mänz, N. (2007). Untersuchung des Suchverhaltens im Web - Interaktion von Internetnutzern mit Suchmaschinen, Dissertation, Verlag Dr. Kovač, Hamburg.

- Schmidt-Mänz, N. (2006): Erkenntnisse aus dem Suchverhalten im Web – "Muster in Suchanfragen", in: Eberspächer, J./ Holtel, St. (Eds): Suchen und Finden im Internet, Springer, Heidelberg.
- Schmidt-Mänz, N. / Bomhardt, Ch. (2005): Wie suchen Onliner im Internet?, in: Science Factory 2/2005, Absatzwirtschaft.
- Schmidt-Mänz, N./ Gaul W. (2005). Web Mining and Online Visibility, in: Weihs, C.; Gaul, W. (Eds.) (2005): Classification – the Ubiquitous Challenge. Springer, Heidelberg, 418-425. Proceedings, Gesellschaft für Klassifikation, 2004.
- Schmidt-Mänz, N./ Koch, M. (2006). A General Classification of (Search) Queries and Terms, Proceedings of the Third International Conference on Information Technology: New Generations (ITNG'06), Las Vegas, Nevada, USA, 375-381.
- Schmidt-Mänz, N./ Koch, M. (2005): Patterns in Search Queries, in: Baier, D., Decker, R., and Schmidt-Thieme, L. (Eds.) (2005): Data Analysis and Decision Support. Springer, Heidelberg, 122-129.

# Prozess- und Ereignisorientierung bei Kunden-Kommunikations-Dienstleistern\*

*Daniel Jobst*

Universität Regensburg  
Institut für Medien-, Informations-  
und Kulturwissenschaft (IMIK)  
Universitätsstr. 31  
93053 Regensburg  
*daniel.jobst@eventprocessing.net*

## **Abstract**

Der vorliegende Artikel beschreibt ein laufendes Promotionsvorhaben am IMIK der Universität Regensburg. Ziel der Arbeit ist es, ein Modell zu entwickeln, welches Konzepte des Prozessmanagements, der Serviceorientierung und der Ereignisorientierung zu einer sinnvollen Systemarchitektur bei Kunden-Kommunikations-Dienstleistern vereint.

## **1 Kunden-Kommunikations-Dienstleister**

Der Begriff Kunden-Kommunikations-Dienstleister (KKDL) dient in der zu Grunde liegenden Arbeit als Allgemeinbegriff für sowohl Firmeninterne wie auch selbständig agierende Dienstleister, die die Kommunikationsschnittstelle großer Unternehmen zu ihren Kunden bereitstellen. An erster Stelle sind hier so genannte „Call Center“ zu nennen, die Kundenanfragen über alle Kanäle hinweg bearbeiten. Kanäle sind zum Beispiel die Telefonie, Fax, E-Mail, SMS oder auch konventionelle Post. KKDL werden dazu genutzt, um eine einheitliche Kundenschnittstelle und Kundenkommunikation bereit zu stellen. Auf Grund einer starken Ausprägung und Integration von Telekommunikationstechnologie in meist heterogene Systemverbünde ergeben sich bei KKDL besondere Anforderungen an die Systemarchitektur, die zum einen dargestellt werden als auch in einen neuen Bezug zu Prozessorientie-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 371-373.

rung, serviceorientierten Architekturen und der Ereignisorientierung gestellt werden.

## **2 Geschäftsprozesse und Dienstorientierung**

Insgesamt wird ein Modell entwickelt, mit dessen Hilfe fachliche Geschäftsprozesse in einer serviceorientierten Architektur (SOA) umgesetzt und automatisiert werden können. Dabei werden spezielle Anforderungen von KKDL berücksichtigt und in das Modell integriert.

Einführend werden zunächst verschiedene Prozess- und Workflowbegriffe definiert und abgegrenzt. Auf Basis der von Prof. Scheer in [Scheer 1991] entwickelten Architektur integrierter Informationssysteme (ARIS) wird ein hierarchisches Prozessmodell weiterentwickelt. Definitionen von SOA werden gegenüber gestellt (vgl. z. B. [Aier & Schönherr 2006]) sowie der betriebswirtschaftliche Nutzen von SOA (vgl. z. B. [Bieberstein et al. 2006]), unter anderem im Hinblick auf KKDL erweitert, dargestellt. Weiterhin wird auf das Auffinden und „Schneiden“ von Services in einer SOA mit Hilfe des erweiterten Prozessmodells eingegangen und grundsätzliche Modellierungskonventionen für die Nutzung des hierarchischen (ARIS)-Prozessmodells als Modellierungssprache für eine SOA aufgestellt.

## **3 Ereignisorientierung**

Das Modell wird im weiteren Verlauf der Arbeit um Komponenten der Ereignisorientierung erweitert. Hier werden zunächst die Konzepte von „Complex Event Processing (CEP)“ (vgl. [Luckham 2003]), „Event Stream Processing (ESP)“ (vgl. [Aggarwal 2006]) und ereignisgetriebener Architekturen (EDA) (vgl. [Mühl, Fiege & Pietzuch 2006]) erläutert.

Ziel ist es, Ereignisse, die in einer EDA mit Hilfe von CEP beziehungsweise ESP verarbeitet werden, bereits bei der Modellierung von Prozessen mit zu berücksichtigen und den Fachanwender (als dem eigentlichen „process owner“) definieren und modellieren zu lassen.

Dazu wird das Modell der ereignisgesteuerten Prozessketten (EPK) (vgl. [Rump 1999]), die zur detaillierte Prozessbeschreibung im hierarchischen Prozessmodell dienen, erweitert, um damit Ereignisquellen in den Prozessen definieren und modellieren zu können. Dies wird außerdem in die Modellierungskonventionen mit aufgenommen.



## **4 Anwendungsbeispiel bei einem KKDL**

Das entwickelte Modell wird anhand eines konkreten Anwendungsszenarios beispielhaft umgesetzt. Als Szenario dienen Anforderungen aus einem realen Projekt im Versandhandel<sup>1</sup>. Darin wird das Prozessmodell auf die fachlichen Prozesse angewandt und um sinnvolle Teile von Kommunikationsprozessen erweitert. (Kommunikationsprozesse dienen grundsätzlich zur Steuerung von Kundenkontakten (z. B. Telefongespräche) während des Prozessablaufs im Call Center.) Die sich daraus ergebende IT-Architektur mit service- und ereignisorientierten Komponenten wird ebenso dargestellt wie ihre grundsätzliche Anwendbarkeit auf andere KKDL.

## **5 Literatur**

- [Aier & Schönherr 2006] Aier, Stephan, Schönherr, Marten (Hg.): Unternehmensarchitekturen und Systemintegration, 2. Auflage, Gito, Berlin 2006.
- [Aggarwal 2007] Aggarwal, Charu (Hg.); Data Streams, Models and Algorithms, Springer, New York et al. 2007.
- [Bieberstein et al. 2006] Bieberstein, Norbert, Bose, Sanjay, Fiammante, Marc, Jones, Keith, Shah, Rawn; Service-oriented architecture compass: business value, planning, and enterprise roadmap, Pearson, Upper Saddle River 2006.
- [Mühl, Fiege & Pietzuch 2006] Mühl, Gero, Fiege, Ludger, Pietzuch, Peter; Distributed Event-Based Systems, Springer, Berlin und Heidelberg 2006.
- [Luckham 2002] Luckham, David; The Power of Events, An Introduction to Complex Event Processing in Distributed Enterprise Systems, Addison-Wesley, Boston et al. 2002.
- [Rump 1999] Rump, Frank, Geschäftsprozeßmanagement auf der Basis ereignisgesteuerter Prozeßketten, Teubner, Stuttgart & Leipzig, 1999.
- [Scheer 1991] Scheer, August-Wilhelm; Architektur integrierter Informationssysteme, Grundlagen der Unternehmensmodellierung, Springer, Berlin et al. 1991.
- [Thieme 2006] Thieme, Jan; Versandhandelsmanagement, Grundlagen, Prozesse und Erfolgsstrategien für die Praxis, 2., aktualisierte und ergänzte Auflage, Gabler, Wiesbaden 2006.

---

<sup>1</sup> Zu allgemeinen Anforderungen, unter anderem zu Prozessen der Bestellannahme im Versandhandel (sowohl telefonisch als auch schriftlich), vgl. [Thieme 2006]. Die konkrete Erlaubnis zur Nennung des Unternehmens und der betroffenen Projektbezeichnungen lagen zum Zeitpunkt der Drucklegung dieses Artikels noch nicht vor.



# Das Rechercherecht – Möglichkeiten und Grenzen\*

*Bernd Juraschko*

Universität Konstanz  
Fachbereich Rechtswissenschaft  
78464 Konstanz  
*Bernd.Juraschko@gmx.de*

## **Zusammenfassung**

Unter dem Begriff Rechercherecht wird die Rechtsmaterie zum Auffinden und Auswählen von Informationen zusammengefasst. Es besteht aus einem materiellrechtlichen und einem verfahrensrechtlichen Teil. Inzwischen spielt neben den Fragen des Zugangs zu den Informationen auch das Haftungsrecht eine immer größer werdende Rolle.

## **Abstract**

Research law means a summery of legal rules for finding and choosing information. It is divided in a procedural law part and substantive law part. Beside the questions how to get the information plays the liability law an increasing important role.

Eine Recherche dient dazu, die gewünschte Information zu finden. Information ist neues Wissen über ein Ereignis, einen Tatbestand oder einen Sachverhalt. Das Rechercherecht ist damit ein Bestandteil des Informationsrechtes. Der Begriff des Rechercherechts ist bisher noch nicht etabliert. Er umfasst die bisher sehr stark zergliederte Rechtsmaterie zum Auffinden und Auswählen von Informationen. Die bisherige allgemeine Einteilung greift in den einzelnen Rechtsgebieten auf sehr unterschiedlichen zu. Beispielsweise ist die Informationsfreiheit grundgesetzlich und damit sehr allgemein geregelt. Dagegen ist das Haftungsrecht über die Spezifikationsebenen: Zivilrecht – Schuldrecht – besonderes Schuldrecht und schließlich Vertragsrecht oder Deliktsrecht erheblich stärker untergliedert. Es ist damit der Gedanke, vom Handlungsgegenstand verwandte Vorschriften, die der herkömmlichen Einteilung Zivilrecht, öffentliches Recht und Strafrecht entstammen, unter

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 375-379.

einem eigenen Rechtsbegriff zusammenzufassen, als Einheit zu verstehen und Beziehungen zwischen den betreffenden Normen herzustellen. Eine solche Einteilung erleichtert bei der teleologischen Auslegung das Normenverständnis und eröffnet neue Gesichtspunkte. Aus den gleichen Gründen sind gesetzlich nicht eigens kodifizierte Rechtsgebiete wie das Bibliotheksrecht entstanden. Auch der Gesetzgeber bedient sich dieser Ordnung, indem er beispielsweise unter dem Begriff des Patentrechts zivilrechtliche, strafrechtliche sowie verfahrensrechtliche Vorschriften in einem Gesetz zusammengefasst hat. Die ständige und erhebliche Zunahme des Wissens und die damit verbundene, ebenfalls in ihrer Bedeutung steigende Bedeutung des notwendigen Auffindens von Informationen und die damit entstehenden rechtlichen Fragestellungen einerseits als auch die sehr starke Zergliederung der betreffenden Vorschriften andererseits sprechen für eine Sammlung unter dem Oberbegriff Rechercherecht.

Nachdem der Grund für eine Einführung eines neuen Oberbegriffs genannt wurde, wird im Folgenden der Inhalt des Rechercherechts exemplarisch beschrieben.

Zunächst ist der Grund, insbesondere die Notwendigkeit einer Recherchetätigkeit festzustellen. So besteht keine Notwendigkeit für eine Recherche, wenn und soweit eine Informationspflicht besteht. So gibt es gesetzliche Informationspflichten. Beispielsweise seien §§ 666 BGB (Auskunfts- und Rechenschaftspflicht des Beauftragten), 16 Abs. 1, 23 Abs. 2, 27 Abs. 2 VGG (Anzeigepflichten beim Versicherungsvertrag) genannt. Darüber hinaus können sich bei besonderen Vertrauensverhältnissen Aufklärungspflichten aus Treu und Glauben ergeben. Diese Fälle stellen jedoch eine Ausnahme dar. Regelmäßig hat sich jeder eigenverantwortlich um seinen Informationsstand zu kümmern. Bei der Recherche handelt es sich daher um eine frei gewählte Tätigkeit; gegebenenfalls um eine Obliegenheit. D.h. niemand wird zu einer Nachforschung gezwungen. Er hat jedoch unter Umständen im Falle einer Versäumnis die Folgen eines solchen Unterlassens zu tragen. Soweit zur „Pflicht“ tätig zu werden.

Nach dem „Ob“, soll nunmehr das „Wie“ angesprochen werden. Die Informationsfreiheit, geregelt in Art. 5 Abs. 1 S. 1 GG, bedeutet, sich frei aus öffentlich zugänglichen Quellen informieren zu können. Zunächst ist eine Unterscheidung zwischen öffentlich-rechtlich und privatrechtlich organisierten Quellen zu treffen.

Gegenüber staatlichen Einrichtungen kann sich die suchende Privatperson auf die Informationsfreiheit gem. Art. 5 Abs. 1 S. 1 GG berufen, soweit es die Quellen „frei zugänglich“ sind.

Gegenüber Privatpersonen z. B. dem Inhaber eines privaten Archivs gibt es keine verfassungsrechtlichen, sondern nur einfach gesetzlich geregelten, zivilrechtlichen Grundlagen. Als Kehrseite der Auskunftsansprüche sind Rechercherechte wie z. B. § 260 BGB zu nennen.

Weiterhin ist zu berücksichtigen, dass sich staatliche Einrichtungen im Gegensatz zu Privatpersonen grundsätzlich nicht auf einen Grundrechtsschutz berufen können. Dies bedeutet, dass sich staatliche Einrichtungen wie öffentlich-rechtlich organisierte Bibliotheken keinen grundgesetzlich abgesicherten Auskunftsanspruch besitzen. Sie bedürfen dagegen eigener Ermächtigungsgrundlagen. Diese befinden sich regelmäßig in einfach-rechtlichen Normen. Soweit zu der Beziehung zwischen Quelleninhaber und Suchendem.

Häufig bleibt es aber nicht bei dem Zweipersonenverhältnis. So kann auf Seite des Suchenden eine weitere Person hinzugezogen werden. Diese kann ebenfalls juristischer oder privatrechtlicher Natur sein. Exemplarisch soll hier die Unterstützung durch eine privatrechtliche Person näher erörtert werden.

Wird eine Recherche von einer anderen unabhängigen Person durchgeführt, als derjenigen, die die Information benötigt, kann ein Recherchevertrag vorliegen. Je nach Willensrichtung der Parteien kommt alternativ ein Gefälligkeitsverhältnis bzw. ein Gefälligkeitsvertrag in Betracht. Ein solches Verhältnis ist regelmäßig dann anzunehmen, wenn die zu gebende Information allgemein im üblichen Gespräch erwartet werden kann. Regelmäßig handelt es sich um schnell und leicht zu ermittelnde Informationen. Ein Indiz, nicht aber ein Beweis, für die reine Gefälligkeitsbehandlung ist die Unentgeltlichkeit. Dagegen wird bei komplexeren, zeitaufwändigeren Recherchen eher ein Recherchevertrag anzunehmen sein. Dabei handelt es sich um einen Geschäftsbesorgungsvertrag. Je nach seiner Ausgestaltung enthält dieser dienst- oder werkvertragliche Elemente. Steht die Tätigkeit ohne Erfolgsgarantie im Mittelpunkt, so herrschen dienstvertragliche Elemente gem. § 611 BGB vor. Wird hingegen ein Erfolg geschuldet, so liegt der Schwerpunkt im Werkvertragsrecht gem. § 631 BGB. Beispielsweise ist der Auftrag zu ermitteln, ob eine technische Entwicklung eine Neuheit im Sinne des Patentrechts ist oder ob sie lediglich dem Stand der Technik entspricht, ein auf einen Erfolg gerichteter Auftrag. Die Unterscheidung ist hinsichtlich der Bindungswirkung und einer eventuellen Haftung von Bedeutung.

Von wachsendem Interesse ist vor allem das Haftungsrecht für Recherchetätigkeiten. Diese Entwicklung geht mit der wachsenden Bedeutung der gewerblichen Schutzrechte und ihrer Vermarktung einher. Hauptanwendungsfälle sind bisher

Klagen gegen Patentanwälte wegen fehlerhafter Neuheitsrecherche und Empfehlung einer aussichtslosen Patentanmeldung. Verkürzt dargestellt geht es neben dem fehlerhaften Ergebnis gem. § 276 BGB um die Verantwortlichkeit des Schuldners. Grundsätzlich hat er Vorsatz und Fahrlässigkeit zu vertreten. Regelmäßig geht es um die Frage eines fahrlässigen Verhaltens, also ob er die im Verkehr erforderliche Sorgfalt außer Acht gelassen hat. An dieser Stelle wird der Ball wieder von den Juristen an die Informationsspezialisten zurückgegeben. Sie haben festzulegen, welche Anforderungen an eine entsprechende Recherche zu stellen sind. Zu prüfen ist danach der fachliche Umgang mit den Quellen in einer angemessenen Zeit, die Nachvollziehbarkeit der gefundenen Ergebnisse und den daraus resultierenden Schlussfolgerungen.

Kein Recht besteht schrankenlos. Daher ist auch der Rahmen für die Recherchetätigkeit aufzuführen.

Seine Grenzen findet die Recherchetätigkeit bei Straftaten gegen den Staat in den §§ 96 ff. StGB. Sowie eine Begrenzung der Ermittlungen gegen Privatpersonen gem. §§ 202, 202a, 203 StGB sowie § 206 StGB und § 19 UWG. Allgemein ausgedrückt handelt es sich hierbei um strafrechtliche Verbote gegen ein Eindringen in Informationen, die geheimgehalten werden und deren Veröffentlichung auf Grund der Art und Weise der Informationsbeschaffung allgemein missbilligt werden. Soweit die Recherche von Seiten der öffentlichen Hand durchgeführt wird, wird ihr Umfang durch die Ermächtigungsgrundlage begrenzt.

Eine Grenze der anderen Art ist das Rechtsberatungsgesetz. Hiernach darf die reine Recherche von nicht autorisierten Personen nicht zu einer kompletten Rechtsberatung ausgebaut werden. Bei dieser Grenze geht es aber nicht um die Recherche an sich, sondern darum, wie sie dargeboten und eventuell angereichert wird.

Soweit zum materiellen Teil des Rechercherechts.

Das Rechercherecht besteht nicht nur aus dem bisher angesprochenen materiellen Recht, sondern auch aus einem verfahrensrechtlichen Teil. Hier soll ein Beispiel aus dem Bereich der Abrechnungen genannt werden: Förderlich für Recherchefreudigkeit ist es, wenn es gelingt, die Kosten auf fremde Schultern verschieben zu können. Im Allgemeinen hat der Auftraggeber die Gebühren bzw. das Entgelt für eine Recherche zu bezahlen. Von Interesse sind die Ausnahmen, da in diesem Fall der Auftraggeber eher geneigt sein wird, weitere Recherchen durchführen zu lassen. Mithin sind solche Ausnahmen geeignet die Nachfrage nach einer qualifizierten Informationsbeschaffung zu erhöhen. Eine solche Ausnahme sieht beispielsweise das OLG Hamm in seinem Beschluss vom 27.12.2001 (Az.: 23 W 469/01). Danach

gehört der Aufwand für eine patentrechtliche Eigenrecherche zu den besonders zu erstattenden Kosten. Im Gegensatz zu der Sichtung, Auswahl und Auswertung gehört die Beschaffung anderer für die rechtliche Beurteilung notwendiger Muster nicht zu der durch § 15 Abs. 5 GebrMG abgedeckten allgemeinen Tätigkeit.

Recherche erforscht nicht nur, sondern sie verbindet auch. So ist an der Universität Konstanz ein neuer Ansatz zur Zusammenarbeit verschiedener Fachbereiche im Gange. Aus dem Fachbereich Rechtswissenschaft heraus ist in Zusammenarbeit mit den naturwissenschaftlichen Fachbereichen und voraussichtlich dem Technologie-Büro Baden-Württemberg eine Novelle der „Patentoffensive“ im Werden. Neben einer Sensibilisierung für die Rechte zum Schutze des geistigen Eigentums steht darauf aufbauend eine Einführung in die Schutzrechtrecherche und eine Betrachtung des Recherchehaftungsrechts auf der Agenda.





# Ausblick auf die modellgetriebene, mustergestützte Sicherheit in serviceorientierten Architekturen\*

*Heiko Klarl<sup>1, 2</sup>*

<sup>1</sup>Universität Regensburg  
Institut für Medien-, Informations-  
und Kulturwissenschaft  
93040 Regensburg

<sup>2</sup>iC Consult GmbH  
Keltenring 14  
82041 Oberhaching  
*h.klarl@klarl.eu*

## **Zusammenfassung**

Mit dem Konzept der serviceorientierten Architektur und der damit erwarteten schnelleren und flexibleren Anpassung von Geschäftsprozessen (GP) werden neue Ansätze zur Absicherung solcher Systeme nötig. Sicherheitsanforderungen werden bisher jedoch meist abseits der Geschäftsprozessmodelle betrachtet und separat definiert. Dies führt zu zwei parallelen Modellen der Anforderungsbeschreibung, d. h. die funktionellen Anforderungen des GP existieren losgelöst von dessen nicht-funktionellen Anforderungen, insbesondere der sicherheitsrelevanten. Regelmäßige Änderungen des GP können so zu Inkonsistenzen bei der Absicherung führen. Im Rahmen eines Promotionsverfahrens sollen Möglichkeiten gefunden werden, auf globalen Richtlinien basierende Sicherheitsanforderungen enger mit den Geschäftsprozessmodellen zu verknüpfen und daraus valide Policies für die Sicherheits-Integrations-Ebene zu erzeugen.

## **Abstract**

With the emerging trend of service-oriented architectures and a faster and more flexible adaption of business processes, new security paradigms are being required. Security requirements are mostly separated from business processes. Thus, functional and non-functional requirements exist in two different models. Changes of the business process could therefore lead to an inconsistency of its security models. The future research work concentrates on linking security requirements – based on global policies – to business process models for generating valid policies for the security integration layer.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informatikwissenschaft. Konstanz: UVK, 381-383.

## I Einleitung

Das Paradigma der serviceorientierten Architektur verspricht der Unternehmens-IT die schnelle und flexible Anpassung von GP, um einerseits auf Veränderungen des Marktes reagieren und andererseits gesetzliche Vorschriften umsetzen und dadurch Compliance-Anforderungen einhalten zu können. Die Absicherung der serviceorientierten Architektur kann über eine Sicherheits-Integrations-Ebene erfolgen und ist somit nicht mehr plattform- oder applikationsorientiert [Nakamur et al. 2002]. Die funktionellen Aspekte eines GP können mit bewährten Methoden modelliert werden (vgl. [Keller et al. 1992], [UML]), die Beschreibung der nichtfunktionellen Sicherheitsanforderungen ist jedoch noch immer weitgehend von diesen Modellen losgelöst und existiert oftmals nur als Spezifikation in verschiedenen Dokumenten (vgl. [Lodderstedt et al. 2002]). Ziel der künftigen Entwicklung muss es sein, die nichtfunktionellen Anforderungen in das Modell zu integrieren und somit eine engere Kopplung zwischen den funktionellen Anforderungen des GP und seinen nichtfunktionellen Sicherheitsanforderungen zu erreichen.

## 2 Vorgehensweise und Aufbau der Arbeit

Nichtfunktionelle Sicherheitsanforderungen eines GP setzen sich aus zwei verschiedenen Bereichen zusammen. Auf der einen Seite existieren globale Richtlinien und Anforderungen (organisatorische Sicherheits-Meta-Policies), die durch die Sicherheits-Strategie des Unternehmens vorgegeben sind, auf der anderen Seite existieren zusätzlich die fachlichen Sicherheitsanforderungen (Geschäfts-Policies) an den jeweiligen GP. Diese können eine Konkretisierung der Sicherheits-Meta-Policies oder rein fachliche Anforderungen repräsentieren. Die zukünftige Forschungsarbeit teilt sich in verschiedene Schritte auf. Im ersten Schritt ist zu klären, inwiefern bestehende Modelle zur Beschreibung der *Enterprise Architecture* (vgl. [TOGAF], [Zachman 2003]) die Abbildung von organisatorischen Sicherheits-Meta-Policies unterstützen bzw. um diese erweitert werden können und welchen sinnvollen Umfang diese Sicherheits-Meta-Policies annehmen können. Der zweite Schritt umfasst vor allem die Belange der fachlichen Sicherheitsanforderungen. Es ist zu untersuchen, wie diese basierend auf den Sicherheits-Meta-Policies durch die Fachseite konkretisiert werden können. Ferner sind Wege zu finden, wie die Fachseite bei der Konkretisierung unterstützt werden kann und wie anschließend die formalisierten Geschäfts-Policies in das Modell des GP eingefügt werden können, um daraus valide Infrastruktur-Policies für die Sicherheits-Integrations-Ebene er-

zeugen zu können. Sind diese Arbeiten fortgeschritten, kann der Einsatz von Business Process Security Pattern (BPSP) als mögliche Unterstützung für die Fachseite evaluiert werden. BPSP, in Anlehnung an die in [Tatsubori et al. 2004] vorgestellten *Idioms*, bündeln für eine Anforderungssituation verschiedene Geschäfts-Policies und lassen sich daher als „Lösungsschablone“ verwenden. Das Auffinden von BPSP wird vermutlich über zwei verschiedene Wege möglich sein. Wiederkehrende Anforderungen, die mehrere Geschäfts-Policies umfassen können, können aufgrund ihrer logischen Zusammenhänge zu BPSP zusammengestellt werden. Als Beispiel kann die Benutzer-Authentifizierung angeführt werden. Kommt der Benutzer aus dem Intranet, so genügt ein vertrauenswürdiger SAML-Token, kommt er allerdings aus dem Internet, so wird eine Zwei-Faktor-Authentifizierung erforderlich. Andere BPSP werden allerdings nicht so einfach aufzufinden sein und erst erfahrungsgestrieben durch Analyse der modellierten Geschäfts-Policies und ihrer Zusammenhänge beschrieben werden können. Ist eine Menge von BPSP identifiziert, muss eine praktische Evaluierung ihren Nutzen für die Fachseite bewerten.

### 3 Literatur

- [Keller et al. 1992] Keller, G.; Nüttgens, M., Scheer, A.: Semantische Prozessmodellierung auf der Grundlage Ereignisgesteuerter Prozessketten (EPK), *Veröffentlichungen des Instituts für Wirtschaftsinformatik (IWi), Universität des Saarlandes* (1992)
- [Lodderstedt et al. 2002] Lodderstedt, T., Basin, D.A., Doser, J.: SecureUML: A UML-Based Modeling Language for Model-Driven Security. In: International Conference on The Unified Modeling Language, S. 426 – 441 (2002)
- [Nakamur et al. 2002] Nakamur, Y., Hada, S., Neyama, R.: Towards the Integration of Web Services Security on Enterprise Environments. In: Symposium on Applications and the Internet (SAINT) Workshops, S. 166 (2002)
- [Tatsubori et al. 2004] Tatsubori, M., Imamura, T., Nakamura, Y.: Best-Practice Patterns and Tool Support for Configuring Secure Web Services Messaging. In: IEEE International Conference on Web Services, S. 244 (2004)
- [TOGAF] The Open Group: The Open Group Architecture Framework, <http://www.opengroup.org/architecture/togaf8-doc/arch/toc.html> (10.04.2007)
- [UML] Object Management Group, Inc.: Unified Modeling Language (UML), <http://www.uml.org/#UML2.0> (2007) (10.04.2007)
- [Zachman 2003] John A. Zachman: The Zachman Framework: A Primer for Enterprise Engineering and Manufacturing, Electronic book published on [www.zachmaninternational.com](http://www.zachmaninternational.com) (2003)



# Virtuelle Mitglieder in virtuellen Teams – Kompensation defizitärer Rollen durch Simulation

## Unterstützung virtueller Teams beim Lernen einer Programmiersprache\*

*Ralph Kölle*

Universität Hildesheim  
Informationswissenschaft  
Marienburger Platz 22  
31141 Hildesheim  
*koelle@uni-hildesheim.de*

### 1 Einleitung

Dem kooperativen Lernen werden gegenüber individuellen Lernformen viele Vorteile bescheinigt. So fördere die Kooperation einen höheren Lernerfolg, Sozialkompetenz und erhöhte Lernmotivation. Auch die „*Entwicklung komplexer Softwaresysteme findet heutzutage meistens arbeitsteilig in – zunehmend räumlich verteilten – Teams statt*“ (Tietze & Schümmer 2001, S.264). Beim Programmieren lernen werden daher theoretische Inhalte durch tutoriell unterstützte Gruppenübungen und -projekte angereichert. Die technische Übertragung dieser Tutorien ins E-Learning erreicht man durch die Kombination kooperativer Editoren und Werkzeugen der computervermittelten Kommunikation wie z. B. Chat. Werden auch Teletutoren, also die Personen, die virtuelle Teams bei ihrer Arbeit unterstützen, durch elektronische Tutoren ersetzt oder ergänzt, so erreicht man für die virtuellen Teams eine absolute Unabhängigkeit von räumlichen und zeitlichen Gegebenheiten (Rautenstrauch 2001).

### 2 Modellbildung

Bei intelligenten tutoriellen Systemen sind solche virtuellen Tutoren wie ihre menschlichen Kollegen in der Lage, sich dem Wissensstand des Lernenden anzu-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 385-389.

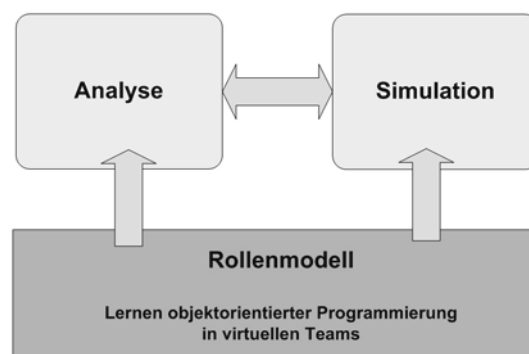
passen. Was beim individuellen Lernen relativ gut funktioniert, lässt sich jedoch nicht direkt auf das kooperative Lernen übertragen, da dort nicht individuelles Wissen, sondern Teamwissen modelliert werden muss.

Das an der Universität Hildesheim entwickelte VitaminL-System modelliert nicht unmittelbar das Teamwissen, sondern bildet sein Teammodell auf Basis eines Rollenmodells (Kölle & Langemeier 2005). Dieses Rollenmodell basiert auf dem Modell von Spencer & Pruss (1995) und besteht aus zehn nach technischen, integrierenden und sozialen Eigenschaften klassifizierten Rollen wie bspw. dem Informationsbeschaffer, dem Berater oder dem Schlichter. Im Gegensatz zu anderen Modellen, die in der Regel davon ausgehen, dass alle Rollen möglichst gleichmäßig besetzt sein sollten, werden die Rollen bei VitaminL zusätzlich nach ihrer Relevanz für den Anwendungsfall der kooperativen Programmierung bewertet und geordnet.

Während der Arbeit des Teams wird dieses auf Ausprägungen einzelner Rollen analysiert. Sollte eine wichtige Rolle unterrepräsentiert sein, übernimmt der virtuelle Tutor diese Rolle, indem er dem Team Hilfsmaßnahmen anbietet, die dieser Rolle zugeordnet sind. Hat das Team bspw. Probleme mit der Syntax oder der Semantik der Programmiersprache, sucht der Informationsbeschaffer geeignete Dateien aus einer Beispielsammlung heraus und präsentiert diese. Er gleicht somit Defizite im Team aus, indem er als virtuelles Mitglied die Funktionen der entsprechenden Rolle simuliert.

### 3 Systementwicklung

Das VitaminL-System ist grundsätzlich in eine Analyse- und eine Simulationskomponente geteilt. Beide Komponenten basieren auf dem Rollenmodell.



Die Schnittstelle zwischen Analyse und Simulation besteht aus einer Case-based-Reasoning-Komponente, die Problemfälle speichert. Die Fallbasis wird durch

Logfiles vergangener Benutzertests und virtueller Tutorien gespeist und kann durch neue Fälle erweitert werden.

Die modulare, objektorientierte Architektur des Systems, kombiniert mit der Aufspaltung des virtuellen Tutors in zehn Rollen, erleichtert die Weiterentwicklung und Evaluation, da die einzelnen Rollen getrennt voneinander parallel in verschiedenen Teilprojekten entwickelt werden können.

Da die Analysekomponente noch nicht fertig gestellt ist, wurde für die Evaluierung der Simulation zunächst die Rolle des Informationsbeschaffers als wichtigste herausgegriffen und prototypisch implementiert. Außerdem stellt diese Rolle durch die Präsentation von Beispielen kombiniert mit kommunikativen Elementen (im Chat) eine Art generalisierbaren Prototyp für andere Rollen dar. Die Aufspaltung der Hilfe nach fachlicher (mittels Präsentation von Dateien) und sozialer Hilfe (mittels Kommunikation im Chat) orientiert sich am *splitting role tutoring*, wie es Kerres (2005, S.173) vorgeschlägt.

## 4 Evaluierung

Die Hilfe des Informationsbeschaffers basiert auf der Präsentation von Programmierbeispielen, die von kommunikativen Phrasen im Chat begleitet werden. In einem zweistufigen Verfahren wurden jeweils Benutzertests durchgeführt, beide Phasen wurden durch Teilnehmer-Fragebögen abgeschlossen.

Alle Benutzertests wurden als Wizard-of-Oz (Woz)-Experiment („hidden operator“) durchgeführt, d.h. das Verhalten des virtuellen Tutors wurde exakt definiert und entsprechende Operationen manuell von einem Bediener nach einem festgelegten Verfahren ausgelöst. Für die Benutzer bleibt das Verfahren transparent, er glaubt mit einem technischen System zu interagieren (vgl. Rapp & Strube 2002, S.661). Da die Simulationskomponente selbst simuliert wurde, kann man auch von *Simulation der Simulation* sprechen.

Als typische Problemsituationen bei Anfängern wurden zunächst Syntaxfehler in den bearbeiteten Dateien gewählt, ein Compiler führte dazu zyklisch einen Syntax-Check durch.

Während in der ersten manuellen Phase die Beispiele intellektuell von einem geschulten Tutor ausgewählt und präsentiert wurden, übernahm diese Aufgabe in der zweiten halbautomatischen Phase eine softwaretechnische Klassifikationskomponente. Die Ergebnisse beider Phasen wurden schließlich miteinander verglichen um

festzustellen, ob sich die Auswahl relevanter Beispieldateien automatisieren ließe, ob also die Simulation des Informationsbeschaffers simulierbar sei.

Die Ergebnisse lassen auf die Tragfähigkeit des rollenbasierten, tutoriellen Konzepts schließen, konnten aber gleichzeitig großes Weiterentwicklungspotenzial aufdecken. Die Bewertungen der Güte der Hilfsmaßnahmen unterschieden sich beim manuellen und beim halbautomatischen Verfahren nur marginal. Dieses Ergebnis überrascht durchaus, offenbar macht die Geschwindigkeit der automatischen Auswahl die Vorteile der intellektuellen Auswahl wett. Immerhin musste ein Tutor aus etwa 70 Dateien die finden, die am besten passt. Allerdings wurden die kommunikativen Fähigkeiten des virtuellen Tutors stark kritisiert. Dieser beherrschte nur einige Phrasen, um die Präsentation der Beispiele zu begleiten. Insbesondere für soziale Rollen sind die kommunikativen Fähigkeiten des integrierten Chatbots zukünftig deutlich auszubauen.

## 5 Referenzen

- [Kerres 2005] Kerres, Michael: Didaktisches Design und eLearning. In: Miller, D. (Hrsg.): E-Learning. Eine multiperspektivische Standortbestimmung. Haupt, 2005, S. 156 – 182
- [Kölle & Langemeier 2005] Kölle, Ralph; Langemeier, Glenn: Rollen in virtuellen Teams – Analyse und Simulation – Arbeitsbericht vom 21.12.2005.  
<http://www.vitaminl.de/downloads/arbeitsbericht2005-12-21.pdf>
- [Rapp & Strube 2002] Rapp, Stefan; Strube, Michael: An Iterative Data Collection Approach for Multimodal Dialogue Systems. In: Rodríguez, M. G.; Araujo, C.P.S. (Hrsg.): Proceedings of the 3rd International Conference on Language Resources and Evaluation, LREC 2002, Las Palmas de Gran Canaria, 29-31 May 2002. Paris: ELRA, 2002, S. 661 – 665
- [Rautenstrauch 2001] Rautenstrauch, Christina: Tele-Tutoren: Qualifizierungsmerkmale einer neu entstehenden Profession. Bertelsmann, 2001
- [Spencer & Pruss 1995] Spencer, John; Pruss, Adrian: Top Teams – Der Königsweg zu mehr Flexibilität, Effizienz und Erfolg im Betrieb. München: Knaur, 1995
- [Tietze & Schümmer 2001] Tietze, Daniel A.; Schümmer, Till: Kooperative Softwareentwicklung. In: Schwabe, G.; Streitz, N.; Unland, R. (Hrsg.): CSCW Kompendium. Springer, 2001, S. 264 – 275



# POSTER

**Datei**

**ISI2007\_Seite389\_PosterKoelle.pdf**



# Quantitative Bewertung von Wiki-Kollaboration in der Industrie

Wiki bei Bosch nach dem 1. Jahr im Einsatz\*

*Alexander Warta*

Robert Bosch GmbH

Diesel Systems

Wernerstr. 51

70442 Stuttgart

*alexander.warta@de.bosch.com*

## **Zusammenfassung**

Wiki-Kollaboration in Unternehmen vollzieht sich unter anderen Rahmenbedingungen als in öffentlichen Wikis - wie z. B. der Wikipedia. So stellt sich in Unternehmen die Frage, wie Wiki-Kollaboration gefördert und gefordert - und damit letzten Endes gemessen - werden kann. Am Beispiel eines Bosch-Wikis soll exemplarisch gezeigt werden, welche kollaborativen Messgrößen aus der „Wiki-Wissensbasis“ ermittelt werden konnten.

## **I Einführung**

Im Oktober 2005 wurde in der Robert Bosch GmbH ein unternehmensweit zugängliches Wiki-System im Testbetrieb installiert.

Zur Nutzung von Wikis in Unternehmen sind Umfragen von Tim Bartel<sup>1</sup> und von Ann Majchrzak, Christian Wagner und Dave Yates verfügbar [Majchrzak, Wagner, Yates 06]. Eine temporäre Analyse des Wiki(pedia)graph stellen Luciana S. Buriol und andere vor [Buriol et al. 06]. Spezielle wissenschaftliche Beiträge zur Wikipedia finden sich im „Wikipedia-Forschungsportal“<sup>2</sup>.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informatikwissenschaft. Konstanz: UVK, 391-395.

<sup>1</sup> Vgl. <http://wikipedistik.de>, Abruf am 08.01.2007

<sup>2</sup> Vgl. <http://de.wikipedia.org/wiki/Wikipedia:Wikipedistik>, Abruf am 08.01.2007

In diesem Beitrag werden die Wiki-Artikel, die im Testbetrieb zwischen dem 12.10.2005 und dem 17.10.2006 im größten der über 50 „Wiki Spaces“ in der Robert Bosch GmbH eingegeben wurden, untersucht. Dabei stellt sich die Frage, welche Kennzahlen ermittelt werden können und aussagekräftig genug sind, um die weitere Entwicklung der Wiki-Kollaboration im Produktivbetrieb zu unterstützen. Die Mitarbeit war freiwillig, wurde durch jeweils dreistündige Schulungen gefördert und von Befragungen begleitet [Warta 07].

## **2 Diskussion der Wiki-Kenngrößen**

Basis der folgenden Ausführungen sind Arbeiten von Rainer Kuhlen und Wolfgang Semar, die Messgrößen für die Informations- und Kommunikationsbereitschaft in elektronischen Foren vorschlagen [Kuhlen 98] und diese für ein kollaboratives Wissensmanagement-System für die Lehre – K<sub>3</sub> – anpassen [Semar 06]. Auf die Unterschiede zwischen K<sub>3</sub> und dem Bosch-Wiki geht Alexander Warta ein [Warta 07].

### **2.1 Allgemeine Kenngrößen**

An 255 Arbeitstagen wurden insgesamt 986 Artikel geschrieben, das ergibt ein arithmetisches Mittel von 3,9 Artikeln pro Arbeitstag. 10 weitere Artikel wurden am Wochenende geschrieben und werden hier nicht berücksichtigt. Der Median beträgt 0 und die Standardabweichung 11, an allein 132 Arbeitstagen wurde kein neuer Wiki-Artikel begonnen.

Es fällt dazu auf, dass es acht Spitzen der Beteiligung gibt, die sich mit vierzehn längeren Zeiträumen abwechseln, in denen gar keine neuen Artikel geschrieben wurden. In den letzten drei Monaten steigt die Beteiligung merklich an.

Beteiligt waren insgesamt 33 Autoren, das arithmetische Mittel ergibt 30,2 Artikel pro Autor. Der Median beträgt wegen einiger extremer Ausreißer 2, die Standardabweichung entsprechend 101,6. Die Autoren haben jeweils zwischen 1 und 561 Artikel geschrieben, allein 14 Autoren haben nur einen Artikel erstellt. 4 Autoren – ein gutes Zehntel – hat über neun Zehntel aller Wiki-Artikel verfasst.

### **2.2 Spezifische Wiki-Kenngrößen**

Der Zusammenhang zwischen der Anzahl der Versionen eines Wiki-Artikels und deren jeweiliger Häufigkeit ist in diesem Fall antiproportional. Als arithmetisches Mittel hat jeder Wiki-Artikel 10,7 Versionen, der Median beträgt 4 und die Stan-

Standardabweichung 25,7. Die häufigste Versionszahl war 282, nur eine Version hatten 133 Artikel.

Was zeichnet einen Wiki-Artikel mit sehr vielen Versionen aus? Entweder ist ein sog. „Edit war“ aufgetreten oder er hat über einen längeren Zeitraum einen hohen Reifegrad erreicht – beides ließe sich auch aus einer geeigneten Visualisierung einzelner Wiki-Artikel ablesen, wie sie beispielsweise Fernanda Viégas, Martin Wattenberg und Kushal Dave vorgestellt haben [Viégas, Wattenberg, Dave 04]. In beiden Fällen sind die entsprechenden Artikel für die Wiki Community relevant – wenn auch aus unterschiedlichen Gründen.

Auch das Verhältnis von Autorenzahl pro Artikel zur Artikelzahl mit eben dieser Autorenzahl ist in diesem Fall antiproportional. Ein Artikel wurde im arithmetischen Mittel von 2,6 Autoren erstellt, der Median beträgt 2 und die Standardabweichung 2,3.

### **2.3 Weitere interessante Kenngrößen**

Die bisher vorgestellten Kenngrößen sollten deutlich machen, wie reich die Wiki-Basis an verwertbaren Daten ist und wie die Auswertung dieser Daten als ein erster Schritt zu einem „Wiki-Reifegrad-Modell“ aussehen kann.

Interessant für die Wiki-Kollaboration ist insbesondere auch die Häufigkeit und Art der Verlinkung (innerhalb des Wikis, aus dem Wiki ins Intra- oder Internet), die bisher noch nicht abschließend untersucht wurde.

Einige Beispiele für weitere interessante kollaborative Kenngrößen sind somit:

- Wie hoch ist der Vernetzungsgrad eines einzelnen Artikels? Wie viele externe und interne Links weist er - in Abhängigkeit zu seiner Länge - auf?
- Wie hoch ist der Interaktionsgrad innerhalb eines einzelnen Artikels?
- Wie viele Versionen wurden von wie vielen Autoren erstellt?
- Wie oft wurde ein Artikel gelesen und wie oft erwuchs daraus ein Kommentar?
- Gibt es einen typischen Lebenszyklus für Wiki-Artikel, an dessen Ende eine „stabile Version“ steht?
- Sind aus der Wiki-Wissensbasis verschiedene Typen von Autoren ablesbar?

## **3 Ausblick**

Die vorgestellten Wiki-Kenngrößen haben deutlich gemacht, dass im betrachteten Space des Bosch-Wikis noch nicht in vollem Umfang von wirklicher Wiki-Kol-

laboration gesprochen werden kann. Die Beteiligung dürfte sich in Zukunft erhöhen, da das Testsystem mittlerweile in die stabilere und performantere Produktivumgebung überführt wurde. Außerdem wurden für die Erprobungsabteilung, die diesen Space nutzt, bestimmte Inhalte definiert, die künftig nur noch innerhalb des Wiki dokumentiert werden sollen. Für die Zukunft ist die Verfolgung der hier vorgestellten und weiterer Kenngrößen im Produktivsystem interessant. Außerdem bieten sich Space übergreifende Vergleiche an.

All diese quantitativen Größen können allerdings nur hinreichender Bestandteil eines erstrebenswerten „Wiki-Reifegrad-Modells“ sein. Notwendig wären darin auch qualitative Bestandteile. In diesem Zusammenhang stellt sich außerdem die Frage, welche Wiki-Daten in welcher Form aus rechtlicher, informationsethischer, aber auch technischer Sicht überhaupt regelmäßig erhoben werden können und sollen.

Der „Wiki Way of Bosch“ beginnt gerade erst, sich auszuprägen.

## 4 Literaturverzeichnis

- [Buriol et al. 06] Buriol, L.S. et al (2006). *Temporal Analysis of the Wikigraph*. Dipartimento di Informatica e Sistemistica, Università di Roma “La Sapienza”, Rom 2006
- [Kuhlen 98] Kuhlen, R. (1998). *Mondlandung des Internet. Elektronische Kommunikationsforen im Bundestagswahlkampf 98*. UVK, Konstanz 1998, 50f.
- [Majchrzak, Wagner, Yates 06] Majchrzak, A; Wagner, C.; Yates, D. (2006). *Corporate Wiki Users: Results of a Survey*. In: WikiSym '06, ACM, Odense 2006
- [Semar 06] Semar, W. (2006). *Evaluation of a Benchmark System for Analyzing Collaborative Group Performance as Part of an Educational Online Knowledge Management System*. In: Proceedings of the International Conference on Information and Knowledge Engineering IKE '06, CSREA Press, Las Vegas 2006, 217–223
- [Viégas, Wattenberg, Dave 04] Viégas, F.B.; Wattenberg, M.; Dave, K. (2004). *Studying Cooperation and Conflict between Authors with History Flow Visualizations*. In: Conference on Human Factors in Computings Systems (CHI) 2004, 6, 575–582.
- [Warta 07] Dittler, U.; Kindt, M.; Schwarz, C. (Hrsg.) (2007). *Wiki-Einführung in der Industrie – Herausforderungen und Chancen am Beispiel von Robert Bosch Diesel Systems*. In: Online Communities als soziale Systeme. Wikis, Weblogs und Social Software im E-Learning. Waxmann, Münster 2007, 41–60

# POSTER

**Datei**

**ISI2007\_Seite395\_PosterWarta.pdf**





# Studentisches Programm der ISI 2007\*

*Hanna Knäusl*

Universität Regensburg  
Institut für Medien-, Informations- und Kulturwissenschaft  
Informationswissenschaft  
93040 Regensburg  
*hanna@knaeusl.de*

## Abstract

Das 10. Internationale Symposium für Informationswissenschaft 2007 in Köln wird durch ein studentisches Programm ergänzt. Studierende haben die Möglichkeit, sich vor einem Fachpublikum, welches sich aus allen Besuchern der ISI zusammensetzt, zu präsentieren. Im Zentrum des studentischen Programms steht der fachliche Austausch der Studierenden von verschiedenen Studienstandorten. Durch die Vielzahl der verschiedenen Themenschwerpunkte an den einzelnen Institutionen kam ein interessantes Vortragsprogramm zustande. Die nachfolgende Zusammenfassung der Vortragsthemen basiert auf den Themenbeschreibungen der Autoren.

## 1 Universitätsverlage in Deutschland – Dienstleister für die Wissenschaft. Zusammenfassung einer Bestandsaufnahme

Heinz Pampel, Hochschule der Medien Stuttgart

Der Autor hat sich im Rahmen seiner Diplomarbeit an der *Hochschule der Medien Stuttgart* mit der Landschaft der deutschen Hochschulverlage befasst und liefert anhand seiner Ergebnisse eine Beschreibung dieses Typs.

Der Autor unterscheidet Hochschulen, die ihre verlegerische Tätigkeit selbst organisieren und solche, die mit Privatverlagen kooperieren und der freizugänglichen Verbreitung ihrer Publikationen zurückhaltend gegenüber stehen. Die meisten der untersuchten Verlage sind als Einrichtungen der Universitätsbibliotheken tätig. Die Mehrheit der Verlage wird von den Hochschulen subventioniert.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 10. Internationales Symposium für Informationswissenschaft. Konstanz: UVK, 397-402.

Als Herausforderungen beschreibt der Autor, wie sich der junge Verlagstyp Universitätsverlag in einem Spannungsfeld zwischen Wissenschaft und Literaturversorgung etablieren muss. Zentraler Akteur, der über die Entwicklung der Universitätsverlage entscheidet ist der Wissenschaftler – als Autor und als Leser.

Eine weitere Herausforderung für die Hochschulverlage ist es, qualitativ hochwertige Dienstleistungen anzubieten. Neben der inhaltlichen und formalen Qualitätssicherung müssen technische Qualitätssicherungsmaßnahmen getroffen werden, um sich im Bereich des elektronischen Publizierens zu profilieren. Auch gegenüber der Kritik der Privatverlage ist eine Positionierung unerlässlich. Dabei sollten die Universitätsverlage ihren Autoren einen sichtbaren Mehrwert bieten – insbesondere dadurch, dass sie den Open Access zu den von ihnen verlegten Publikationen gewährleisten.

Für Hochschulen und ihre Bibliotheken können Universitätsverlage ein Element bei der Etablierung alternativer Publikationsstrukturen sein, um die an der Hochschule gewonnenen Erkenntnisse getreu dem Ausspruch des Gründers der *Johns Hopkins University Press* (ältester heute noch existierender Universitätsverlag in den USA), Daniel Coit Gilman, aus dem Jahr 1878 zu verwirklichen:

It is one of the noblest duties of a university to advance knowledge, and to diffuse it not merely among those who can attend the daily lectures – but far and wide. (Daniel Coit Gilman)<sup>1</sup>

## Literatur

- [Givler 2002] Givler, Peter: University Press Publishing in the United States. In: Abel, Richard E. u a. (Hrsg.): *Scholarly Publishing. Books, Journals, Publishers and Libraries in the Twentieth Century*. New York : Wiley, 2002, S. 107-120.
- [Hawes 1967] Hawes, Gene R.: *To advance knowledge. A Handbook on American University Press Publishing*. New York : American University Press Services, 1967.
- [Pampel 2007] Pampel, Heinz: *Universitätsverlage im Spannungsfeld zwischen Wissenschaft und Literaturversorgung. Eine kritische Bestandsaufnahme*. Frankfurt am Main : Fachhochschulverlag, 2007.

## 2 Swarm Intelligence

Olivia Macoli, Frederike Schleinitz, Fachhochschule Köln

Die Autorinnen beschreiben das Ziel, die Idee der "Swarm Intelligence" zu realisieren, was zu diesem Zweck bereits unternommen wurde, und welche Pläne es für die

---

<sup>1</sup> Zitiert nach [Hawes 1967], S. 1.

Zukunft gibt. Der Begriff "Swarm Intelligence" wird verwendet im Zusammenhang mit Ameisen oder Bienen Schwärmen. Diese Beispiele zeigen, dass Gruppen intelligenter sind als die einzelnen Individuen, aus denen sie sich zusammensetzen. Diese spezielle Gruppen Intelligenz ist das Ergebnis aus dem Zusammenwirken aller Teile einer Gruppe.

Als besondere Merkmale der "Swarm Intelligence" beschreiben die Autorinnen: Flexibilität, Robustheit und Selbstorganisation<sup>2</sup>.

Auf Grund der Entwicklung der Informations- und Kommunikationstechnologien verändert sich auch das Kommunikationsverhalten zwischen den Individuen. Firmen werden immer mehr aus kleinen Teilen zusammengesetzt sein. Diese bilden eine Art Schwarm, der dann die Ansprüche der Kunden besser befriedigen kann. Die Autorinnen zitieren A. Picot, der die Entwicklung zur "virtuellen Firma" beschreibt<sup>3</sup>. Es lassen sich zwei grundlegende Tendenzen aufzeigen:

- "Managing by wire": Neue Möglichkeiten für das Controlling und die Organisation
- "From Market Place to Market Space".

Als Ziel beschreiben die Autorinnen, ein Netzwerk zu schaffen, in dem Studenten Arbeiten veröffentlichen und diskutieren und in Teamwork an Projekten arbeiten können. Es soll eine Kommunikation zwischen Studenten mit ähnlichen Zielen und Ideen aufgebaut werden um die Synergien des Teamwork zu nutzen. Die Arbeit an dem Projekt soll im web stattfinden um Kommunikation und Arbeitsteilung zu unterstützen.

### 3 Informetrie als Methode zur Beschreibung von Technologietrends

Axel Oepkemeier, Fachhochschule Köln

In seiner Diplomarbeit beschäftigt sich der Autor mit einer empirischen Überprüfung des Konzepts des *Hype Cycles*. Dazu verwendet er das *Instrumentarium* der Informetrie bzw. der Bibliometrie und führt Abfragen bei DIALOG durch.

---

<sup>2</sup> Vgl. Harvard Business manager, 6/2001, S. 44.

<sup>3</sup> [http://www.iuk.bwl.uni-muenchen.de/forschung/veroeffentlichungen/picot/vortraege/landesbank\\_manuskript.pdf](http://www.iuk.bwl.uni-muenchen.de/forschung/veroeffentlichungen/picot/vortraege/landesbank_manuskript.pdf), (10.11.2006).

Der Autor überprüft das Konzept anhand von 5 Technologiefeldern aus der TK – Branche: Mobile Messaging (MMS), UMTS (Mobilfunk der 3ten Generation), WiMAX (neuer Breitband-Standard), Mobile Gaming (Handyspiele) und *location based-Services* (LBS, ortsbezogene Mobilfunkdienste). Für diese Technologiefelder hat der Autor Themenzyklen gemessen.

Als neuen Aspekt entwickelt der Autor ein Konzept, wie er mittels Retrieval auf DIALOG nicht nur Berichterstattungsmengen sondern auch die Tendenz der Berichterstattung messen kann. d.h. inwieweit über die Technologien positiv oder negativ berichtet wird.

## 4 Internationales GUI – Design

Ann Eckert, Mareike Küsell, Teresa Märkl, Universität Hildesheim

Die Autorinnen befassen sich in ihrer Arbeit zu den unterschiedlichen kulturellen Ansprüchen an ein Graphisches User Interface mit drei Themengebieten:

- Lokalisierung und Internationalisierung
- Offshoring
- Remote Usability Testing.

Zunächst grenzen die Autorinnen die Begriffe Lokalisierung und Internationalisierung voneinander ab und beschäftigen sich mit der Rolle der *corporate identity* in diesem Zusammenhang. Im Zusammenhang mit dem Thema *offshoring* beschäftigen sich die Autorinnen mit dem Entwicklungsbereich Usability anhand eines Beispiels des *offshoring* in den Asiatischen Raum. Im dritten Themenbereich gehen die Autorinnen auf das Thema der Evaluation mit zeitlich und räumlich voneinander unabhängigen Testleitern und Probanden ein und beschreiben Problematiken, die bei international durchgeführten *Remote Usability Tests* auftreten können.

## 5 inblot – intelligent blog bot Perspektiven für Web 2.0-Mashups

Doris Maassen, FHTW Berlin

Ausgehend von den Ergebnissen ihrer Diplomarbeit "inblot – intelligent blog bot" diskutiert die Autorin die Perspektiven von Web 2.0-Mashups.

In ihrer Diplomarbeit hat die Autorin ein automatisches Kommentierungs-System für Weblogs entwickelt und implementiert, das dem Mashup-Prinzip folgt. Ein Bot,

genannt "inblot", das aktuelle Weblog-Einträge aus deutsch- und englischsprachigen Weblogs analysiert und dort möglichst intelligente Kommentare veröffentlicht, die passende Links enthalten.

Der „inblot“ verwendet das Web als Datenquelle und gibt ein Remix der gefundenen Inhalte zurück. Für die Suche, für die Extraktion von Begriffen zur Charakterisierung eines Weblog-Eintrags und für die Stammformenreduktion von Suchbegriffen werden Webservices eingesetzt. Zur Generierung von Texten werden verschiedene pragmatische Methoden verwendet. Im Zentrum stehen templatebasierte Ansätze sowie Verfahren einer zufallsbasierten Textrekonstruktion. Gleichzeitig erforscht das Projekt Technologien zur Abwehr von Kommentar-Spam. Das System wird komponentenbasiert in objektorientiertem Perl entwickelt.

Ziel der Arbeit ist nicht die massenhafte Veröffentlichung von Kommentaren, sondern die Entwicklung und Implementierung einer Methode zur Generierung und Veröffentlichung von Weblog-Komentaren mit Hilfe von heterogenen Datenquellen und pragmatischen Verfahren der Sprachverarbeitung. In der Arbeit werden die dazu benötigten Technologien wie Weblog-Management-Systeme, Newsfeeds, Webservices und Spamschutzverfahren analysiert. Im nächsten Schritt werden die rechtlichen und ökonomischen Rahmenbedingungen beleuchtet. Wie keine Art von Anwendungen zuvor sind Web-2.0-Mashups von Diensten abhängig, über die sie keinerlei Kontrolle haben. Bei der Entwicklung des „inblot“ wird deshalb versucht, lokale Verfahren anzubieten, die einen Webservice bei Bedarf ersetzen können.

## 5.1 Literatur

- [Maa07] Maassen, Doris: *inblot – intelligent blog bot*. Diplomarbeit an der Fachhochschule für Technik und Wirtschaft Berlin im Studiengang Internationale Medieninformatik.
- [Mus07] Musser, John: ProgrammableWeb: Mashups and the Web as Platform.  
<http://www.programmableweb.com/>, Zugriff: 03.04.2007.
- [War06] Wartala, Ramon: Mashup: Eine Revolution in Zeiten des Web 2.0. In: iX 7/2006.
- [Yaho07] Yahoo Pipes. <http://pipes.yahoo.com/pipes/docs>, Zugriff: 03.04.2007

## 6 Chancen und Risiken von *creative commons*-Lizenzen für den Stakeholderdialog

Nicole Ebber, Britta Jerichow, Fachhochschule Köln

Free Content Lizenzen wie die *creative commons*-Lizenzen (CCL) sollen Urheberinnen ermöglichen, ihre Rechte flexibel zu handhaben und Nutzern den Zugang zu

Informationen erleichtern. Mithilfe der flexiblen Rechtegestaltung und der expliziten Zuweisung von Nutzungsfreiheiten können sie eine hohe Rechtssicherheit auf beiden Seiten gewährleisten.

Basierend auf den Erkenntnissen der Diplomarbeit „*creative commons*-Lizenzen: Chancen und Risiken für das Management von Non-Governmental Organisations“ (Nicole Ebber) werden verschiedene Anwendungsmöglichkeiten der CCL im Dialog einer Non-Governmental Organisation (NGO) mit ihren Stakeholdern vorgestellt.

Als Stakeholder lassen sich bspw. Mitarbeiter, Managerinnen, Lieferanten, Öffentlichkeit, Kooperationspartnerinnen, Konkurrenten oder Medien bezeichnen. Chancen der CCL schlagen sich bspw. im erhöhten Verbreitungsgrad der Werke und damit der erhöhten Reputation der Urheberin, der Zuweisung standardisierter Metadaten, der rechtlich abgesicherten Einbindung externer Ressourcen oder der positiven Netzwerkeffekte nieder. Sie kommen vor allem dann zum Tragen, wenn die Stakeholder ein gewisses Maß an CC-Affinität vorweisen und der Einsatz dieser Lizenzen konsistent mit den inhaltlichen Werten der Organisation ist. Risiken lassen sich vor allem im gesteigerten Personal- und Verwaltungsaufwand bei der Einführung der Lizenzen sowie in der Angst vor Lizenzverletzung und missbräuchlicher Verwendung ausmachen. Darüber hinaus werden die Inkompatibilität mit anderen freien Lizenzen, die fehlende Sicherstellung von Authentizität und Integrität sowie die Förderung der Vermüllungsproblematik kritisiert.

Neben den untersuchten Möglichkeiten im NGO-Bereich bieten die CCL als Instrumente der Kollaboration und Partizipation weitere Potentiale für Open Innovation. Durch ihre Offenheit und Transparenz schaffen sie eine Grundlage für Innovationen bzw. Kreativität und darüber hinaus neue Perspektiven im Kontext von Information und Wissen. Grundsätzlich ist es für den Erfolg der CCL immens wichtig, eine kritische Masse zu erreichen und das allgemeine Bekanntheits- und Akzeptanzlevel der Lizenzen weiter zu erhöhen. Sorgfältige Aufklärungs- und Öffentlichkeitsarbeit soll einen Bewusstseinswandel in Richtung Wissensgesellschaft einleiten, Hinweise für einen erfolgreichen Einsatz der CCL geben und damit immer mehr Menschen von deren Vorteilen überzeugen. Zusammenfassend gelangt die Untersuchung zu der Erkenntnis, dass die Chancen der Lizenzen gegenüber den Risiken überwiegen. Allerdings müssen Faktoren wie die CC-Affinität und der Aufwand bei der Einführung stets in die differenzierte Betrachtung einbezogen werden.

# intelligentCAPTURE und dandelon.com: Collaborative Catalog Enrichment\*

*Manfred Hauer*

AGI – Information Management Consultants  
Neustadt/Weinstrasse, Deutschland  
*manfred.hauer@agi-imc.de, <http://www.dandelon.com>*

## **Zusammenfassung**

Wissenschaftliche Bibliotheken können insbesondere papierbasierte Medien durch Digitalisierung, OCR, maschinelle Indexierung, Dokumentenanalyse, Informationsextraktion und modernes Information Retrieval zurückholen in den Wahrnehmungsbereich ihrer Klientel. Die Indexierung von Bibliothekskatalogen wird anhand von drei Messreihen problematisiert und in einem Retrieval-Test dandelon.com und Google Scholar gegenübergestellt. Dabei fallen OPACs ohne Catalog Enrichment, sprich Inhaltsverzeichnisse, maschinell generierte Deskriptoren und semantisches Retrieval, hinter die neuen Ansätze deutlich zurück. Durch Grundsatzentscheidungen bei Bibliotheksverbünden wird eine starke Expansion derartiger Inhalte vorhergesagt.

Da viele Bibliotheken die gleichen Medien sammeln, bietet sich ein kollaborativer Ansatz an. Im deutschsprachigen Raum ging der Anstoß in diese Richtung nicht unwesentlich von AGI und ihrem Programm intelligentCAPTURE als Software zu Erfassung, Aufbereitung, Konvertierung, Indexierung und Verteilung und dem wissenschaftlichen Suchdienst "dandelon.com" aus. Bei der Produktion werden zusätzlich Collaboration Tools wie Mail, Chat, IP-Telefonie, Application Sharing, Blog und Webservices genutzt.

## **I     Digitales Inhaltsverzeichnis als Brücke zum Buch**

Artikel und Bücher sollten möglichst komplett online sein, sonst werden sie von den heutigen Studenten und zukünftigen Wissenschaftlern kaum noch wahrge-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 403-412.

nommen, gelesen und zitiert. Durch den Medienträgerwandel von Papier zu digitalem Dokument droht eine erhebliche Menge bisher gesammelten Wissens zu versanden. Diese Lücke betrifft insbesondere eine Zeitspanne von ca. 68 Jahren: Medien älter als 70 Jahre sind meist urheberrechtsfrei, also digitalisierbar ohne juristische Klärung und nur die letzten 2-5 Jahre sind von einigen Verlagen bisher zusätzlich online verfügbar, zumindest bei den ganz großen Verlagen.

Etablierte Reputationsverfahren (Impact Factor) hinken hinter den technischen Möglichkeiten her und präferieren klassisch gedruckte Medien. Zitationslisten widerspiegeln nur noch begrenzt das tatsächliche Lese- und Rezeptionsverhalten der Autoren. Gewiss, der Lesekomfort und die Rezeptionsintensität spricht auch weiterhin für gedruckte Medien, doch der zunehmende Wahrnehmungsverlust nicht. Neues Wissen ist nicht weniger betroffen. Dies gilt auch für jeden Artikel in diesem gedruckten Tagungsband.

Nicht ausreichende Budgets zum Kauf digitalen Contents stehen dem entgegen und mittels des Urheberrechts schützen sich Verlage vor rigoroser Digitalisierung ohne Entschädigung.

Ein Mittelweg ist die digitale Erschließung von Inhaltsverzeichnissen, urheberrechtsfrei, weit aussagekräftiger als allein die intellektuelle Erschließung klassischer Bibliothekskataloge, räumlich nicht beschränkt: digital finden, analog lesen.

Als AGI 1996 erstmals einen Buchhandelskatalog mit Inhaltsverzeichnissen erweiterte, war die Idee noch jung. Doch die praktische Umsetzung im großen Stil ist für derzeitige wissenschaftliche Bibliotheken im Jahr 2007 noch immer die Ausnahme. Doch die Lage ändert sich derzeit rasch. Ein gemeinsamer Beschluss der Bibliotheksverbünde in Deutschland und Österreich brach das Eis und die Deutsche Nationalbibliothek plant Aktivitäten. AGI ist in dieser Entwicklung nicht nur Software-Entwickler und Content-Provider, sondern mittlerweile ein führender Dienstleister mit eigenen, dezentralen Scan-Teams.

## **2 Methode: Zusammenspiel von Menschen, Organisationen, Programmen und Servern**

Es entwickelte sich oberhalb des Bodensees eine erste Community von intelligent-CAPTURE-Anwendern, die seit Sommer 2003 ihren Content nicht nur in die lokalen Bibliothekssysteme integrieren, sondern über „dandelon.com“ zusätzlich zentral speichern und gezielt dezentral distribuieren. Dandelon.com ist mehr als ein Dokumenten-Management-System, für die Öffentlichkeit erscheint es primär als



Suchmaschine für wissenschaftliche Literatur mit einer semantischen, crosslingualen Suche.

Mittels Scanning – von bisher rund 1 Million Papierseiten – über Fujitsu-Flachbettscanner (Weltmarktführer bei Scannern mit 40 % Marktanteil – [http://www.fujitsu.com/us/news/pr/fcpa\\_20061213-01.html](http://www.fujitsu.com/us/news/pr/fcpa_20061213-01.html)) wurden bisher rund 280.000 Bücherinhaltsverzeichnisse in 35 Sprachen in bisher 13 Bibliotheken in vier deutschsprachigen Staaten gescannt und auch ein wenig in Italien. Für Ende 2007 prognostiziert der Autor eine Menge von 500.000 Inhaltsverzeichnissen. Die Mehrheit ist in dandelon.com such- und tauschbar. Dazu kommen 470.000 Aufsatztitel. Noch wenige zeigen den kompletten Volltext (vor allem IWP – Informationswissenschaft- und Praxis der DGI), einige zeigen ihn nur in der Vorarlberg Landesbibliothek. Hier greift das Digital Rights Management der Zeitschriftenagentur Swets. Beim Rest führt der Standortnachweis zum jeweiligen Regal. Die gesamte Lösung ist ein komplexes Zusammenspiel von

- immer mehr Bibliotheken, Bibliothekssystemen und Verbundzentren
- mehreren Entwicklungs-, Hostings- und Projektstandorten bisher vorwiegend in Deutschland und Indien – eine virtualisierte Struktur
- der Integration von Thesauri von den Vereinten Nationen, der Europäischen Union, mehreren Dokumentationszentren, Projekten und aus eigener Produktion,
- von Basisprogrammen verschiedener Hersteller, vor allem
  - IBM mit Lotus Notes & Domino als zentrale multimediale Datenbankumgebung und Entwicklungsplattform für Workflows und mit Lotus Sametime für die realtime Collaboration und darin der GTR als Retrieval-Kernel
  - Abbyy's OCR mit zwei FineReader-Versionen (mobile und Engine)
  - Adobe mit Acrobat für Formatierung und Bearbeitung
  - IAI mit der maschinellen Indexierung, genannt CAI- Engine (=Autindex),
  - Z39.50-Client und XML-Konnektoren,
  - Kofax Image Controls für bestmögliche Images und schnelles Scanning, ISIS-Scanner-Treibern für Fujitsu-Scanner (andere möglich), TWAIN-Treiber werden noch unterstützt
- Servern von SUN und Cients unter Microsofts Windows
- Programmierung und Zusammenstellung von AGI.

14 Thesaurus-Datenbanken mit 1,6 Millionen Begriffen, mehrere Medien-Datenbanken, eine Image-Datenbank für Cover-Pages, Logging und mehrere kleinere Datenbanken für Konfiguration, Monitoring, Buchkaufabwicklung, Kundenverwal-

tung etc, gehören zusammen. Mittlerweile ca. 11 Millionen Dokumente / Datensätze insgesamt.

Social Software wird als Zusammenspiel beliebiger Menschen mit zumeist einfachen Editoren verstanden (z. B. Wikis, Weblogs). Bei dandelon.com spielen ausgewählte Menschen – nur Information Professionals, welche Medien gezielt selektieren und Urheberrechte achten – und zahlreiche wissensbasierte Programme, die automatisch über Workflows miteinander interagieren, zusammen. Der Aufwand zum Editieren und Kommunizieren durch Menschen ist gering und hoch zwischen Computern und Programmen.

### **3 Benchmarking: Indexierung- und Retrievaltests**

Viel Aufwand, aber lohnt sich dieser? Studenten und Wissenschaftler greifen prozentual, wenn sie etwas suchen nur noch selten auf Bibliothekskataloge zu – um die 3 % laut DNB. Andererseits zeigen die Anzahl von Fragen, die Bibliothekskataloge monatlich beantworten, dass sie nahe bei kommerziellen Informationsanbietern wie FIZ Technik, STN, GENIOS liegen – also rund 20.000 bis 80.000 Abfragen pro Monat, die Allermeisten davon beziehen sich auf Sachthemen, die Minderheit auf Namen und bibliografischen Angaben. Auch wenn insbesondere Google in Europa den Suchmarkt zu monopolisieren versucht und wir alle helfen jeden Tag mit. Google hat den Suchraum eher erweitert, klassische und auch mehrere neue Angebote aber noch nicht ganz ersetzt, der Angriff läuft. Es bleiben Nischen, dazu zählen die OPACs der wissenschaftlichen Bibliotheken und die verschiedenen, spezialisierten Fachinformationsdatenbanken. Mit drei Messreihen hat der Autor, die Situation beleuchtet:

#### **3.1 Varianz von Suchergebnissen durch grammatische Varianten, Synonyme und Übersetzungen**

Eine Gruppe von 20 Nachdiplomstudenten der HTW in Chur, teils Mitarbeiter der ETH-Bibliothek und anderer Bibliotheken in der Schweiz, alle mit Erfahrungen im Umgang mit dem Schweizer Bibliotheksverbundkatalog NEBIS, sollten im Sommer 2006 Begriffsvarianten und die Trefferzahlen notieren für Google Scholar, zu der Zeit vermutlich um die 7,5 Millionen Dokumente, NEBIS mit ca. 4,5 Millionen Katalogdatensätzen auf Basis von ALEPH und mit Retrieval-Unterstützung durch OSIRIS und dandelon.com mit damals 130.000 Dokumenten (siehe Abb. 1).

Varianz von Suchergebnissen		2. Juni 2006 - NDL-Studium HTW Chur, 20 Studenten			
	Google ca. 7.2 Mio	Nebis 4,5 Mio	Faktor x 1,6	dandelon 130.000 Dokumente	Faktor 54
Suchwort	Scholar	Nebis	x 1,6	dandelon	x 54
prävention	26100	2581	4130	1413	76303
praevention	9690	2581	4130	1283	69282
prevention	1290000	9714	15542	1348	72792
Vorsorge	11400	443	709	10263	554202
Unfall	32700	1434	2294	911	49194
Unfälle	6210	3737	5979	911	49194
Unfaelle	6210	3737	6979	911	49194
accident	616100	4287	6859	642	34668
accidents	435000	8204	13126	142	7668
injury	2020000	503	804	2118	114372
injuries	828000	1280	2048	104	5616
blume	81700	748	1197	661	35694
flower	415000	1518	2429	793	42822
fleur	17000	199	318	786	42444
Blumen	22600	1368	2189	910	49140
flowers	462000	1571	2514	101	5454
fleurs	18800	1297	2075	21	1134
museum	883000	35143	56229	2340	126360
museums	159000	7713	12341	2461	132894
museen	11800	6594	10550	2340	126360
musée	31300	12195	19512	2055	110970
musée	23800	12195	19512	2055	110970
musseum	46	0	0	2152	116208
Territorialisierung	174	3	5	21	1139
Staatsbildung	684	109	174	84	4536
Territorialherrschaft	59	2	3	14	756
Landesherrschaft	327	23	27	63	3402
Landesherr	466	10	16	79	4566
Territorialherr	25	0	0	7	378
Landesherrlich	17	0	0	3	162
"state-building"	22100	762		19	1026
"emergence of territories"	4	0	0	0	0

Abbildung 1

Zum Vergleich wurde in den markierten Spalten die Ergebniszahl von NEBIS und dandelon.com auf die Menge von Google Scholar hochgerechnet, davor die Absolutzahlen. Anbei ein Ausschnitt aus dem Anfang der Messreihe. Sie zeigt in der Spalte, dass z. B. Google Scholar beim Begriff „Blume“ zwischen 17.000 und 462.000 schwankt, NEBIS zwischen 318 und 2500 und dandelon.com zwischen 1100 und 49.000 schwankt, aber insgesamt weit weniger streut, immer wenn, die Varianten in den Thesauri erkannt werden. Geringfügige Varianten eines Suchbegriffs führen zu teils extrem verschiedenen Ergebnissen sowohl innerhalb des jeweiligen Suchsystems als auch zwischen den Suchsystemen. Addiert man die drei Spal-

ten zusammen, so fällt NEBIS als Repräsentant von Bibliothekskatalogen hinter dandelon.com und Google Scholar zurück.

Man kann einwenden, diese Testreihe ist zu kurz, die Begriffsauswahl durch die Studenten zu zufällig, nicht repräsentativ, die Dokumentenmenge von dandelon.com durch die Hochrechnung um den Faktor 54 evtl. falsch bewertet und zu ungenau. Diese Einwände sind richtig. Dennoch, die komplette Messreihe zeigt noch deutlicher die Schwächen des Bibliothekssystems im Vergleich zu Google Scholar, lässt aber im Vergleich von dandelon.com und Scholar keine Aussage zu, die eindeutig für oder gegen dandelon.com spricht, da 130.000 Dokumente angesichts der Anzahl möglicher Themen noch keine statistisch validen Angaben erlaubt. Dennoch, Google liegt wohl keineswegs immer vorne.

Die Messung überraschte die Studenten sehr, solch große Unterschiede waren niemandem bewusst, da man sich meist mit einer Frage und einer Antwort bei den meisten Suchsystemen zufrieden gibt.

Ziel eines Information Retrieval-Systems muss es sein, derartige Schwankungen beim gleichen Begriff möglichst gering zu halten.

### **3.2 Indexierungstest: Varianz von Autor, Indexierer, Leser**

Ist es nötig, so viele Sprachvarianten zu testen und zu vergleichen? Ist unsere Sprache denn nicht hinreichend klar um eine Sache zu beschreiben? Üblicherweise werden Indexierungstests zwischen zwei menschlichen Indexierern oder zwischen Mensch und maschineller Indexierung gemacht. Zur Verbesserung der Indexierungsqualität sind solche Messungen nützlich, sie greifen aber zu kurz, wenn es um Retrieval geht – die Frage muss vielmehr lauten: Kommen Autor und Suchender/Leser zusammen?

Eine Messreihe (Hauer 2005) mit 99 Studenten und 33 Aufsätzen zeigte, dass nur 21 % der zur Beschreibung verwendeten freien Indexierungsbegriffe zwischen drei Personen übereinstimmten. Drei steht für das Wechselspiel von Autor / Indexierer / Sucher. Die Brücke von Autor zu Sucher ist also schwer zu schlagen. Autoren orientieren sich genauso wenig an Klassifikationssystemen wie ihre Leser und damit stehen Bibliothekare recht einsam auf weiter begrifflicher Flur.

### 3.3 Known-Item-Test: Genau das richtige Buch finden

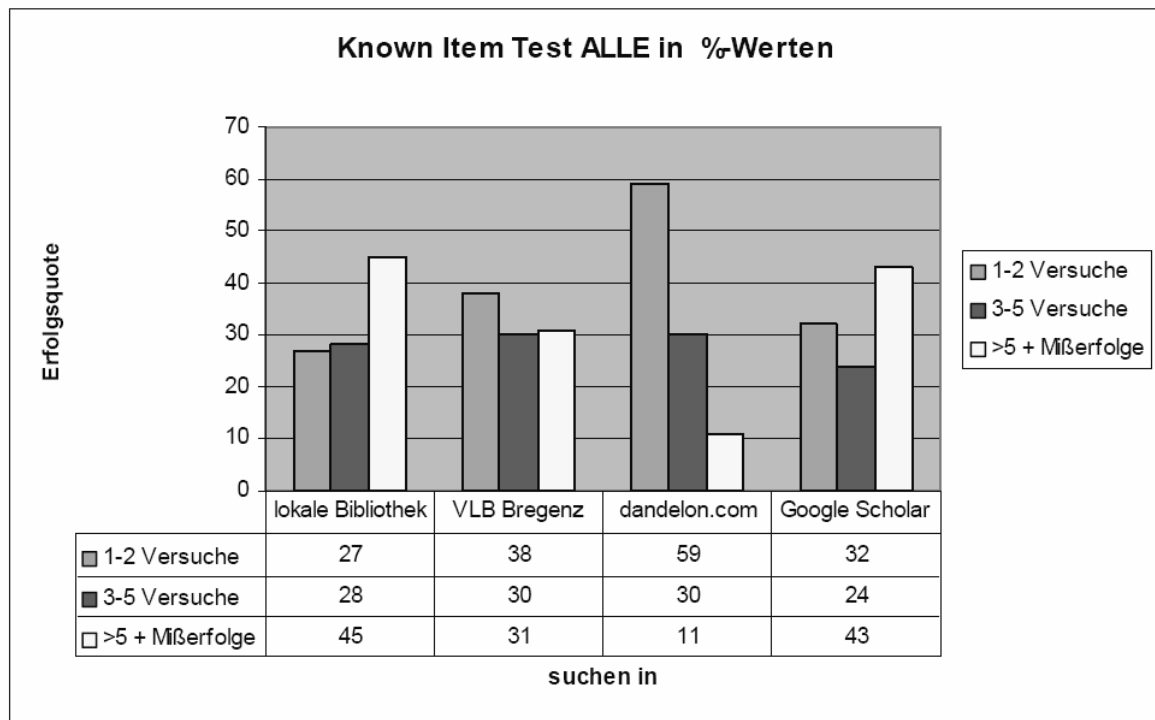


Abbildung 2

Ob dies gelingt, war Gegenstand der dritten hier vorgestellten Messreihe (Hauer: Vergleich der Retrievalleistungen von Bibliothekskatalogen gegen erweiterte und neue Konzepte. In: ABI-Technik Dez. 2005): 75 Studenten mussten 295 Bücher finden, deren Inhalt sie gut kannten, aber nichts über bibliografische Details wussten. Nur durch thematische Recherche konnten die Titel gefunden werden. Die Abbildung 2 zeigt einen Vergleich von lokalen Katalogen (FH Burgenland – 14.000 Titel, HTW-Chur 4000 Titel, NEBIS ca. 4 Mio. Titel), der Katalog der Vorarlberger Landesbibliothek (VLB), welche intelligentCAPTURE-Indexierung und Inhaltsverzeichnisse in den ALEPH-Katalog integriert, dandelon.com und Google Scholar.

Die Grafik zeigt, dass in Bibliothekskatalogen insgesamt 45 % der Studenten scheiterten – dies stimmt nicht für den sehr kleinen Katalog für Chur, während im NEBIS-Katalog die Erfolgsrate extrem niedrig war. Der VLB-Katalog konnte durch das Catalog-Enrichment via intelligentCAPTURE-Deskriptoren Google Scholar bereits übertreffen. Mit 59 % Erfolg mit 1 bis 2 Versuchen lag dandelon.com deutlich vorne. Damals, heute hat dandelon.com nicht mehr 60.000 Titel, sondern über 210.000 öffentlich online, Google Scholar ist gewachsen, eine Wiederholung der Messung wäre spannend.

Scanning von Inhaltsverzeichnissen, die Übertragung von maschinellen Indexaten und PDFs der Inhaltsverzeichnisse an Bibliothekskataloge zahlt sich aus und in Kombination mit semantischen Ressourcen und Ranking-Verfahren führt es signifikant schneller auf jene Medien hin, welche der Anfrager meint.

## **4 Herausforderungen, Hindernisse und Barrieren**

Bezogen auf Inhaltsverzeichnisse gibt es keine wesentlichen Hindernisse durch das Urheberrecht oder Digital Rights Management-Systeme.

Die Kosten für Personalkosten und die Subskriptionsgebühren für die Nutzung von intelligentCAPTURE und dandelon.com werden häufig als Hindernis von Bibliotheken bezeichnet – es ist eher eine Frage der Präferenzen, denn der Aufwand pro Buch kann im Schnitt mit 1,50 € gerechnet werden und dem stehen Einspareffekte bei überflüssigen Ausleihen entgegen. Nur 1 von 10 aus Magazinen ausgeliehenen Titeln wird wirklich gelesen, schätzen viele Bibliothekare. Bei oft entliehenen Titeln sind Mehrfachexemplare nötig – alles nicht zum Nulltarif, aber versteckt in alten Haushaltsposten. Demgegenüber „shareit“ dandelon.com und auch die Verbünde teilen digitalisierte Daten ohne wesentliche Mehrkosten, die Kosten sind also deutlich degressiv mit einer wachsenden Menge von Produzenten.

Hindernisse sind auch persönliche Einstellungen, Ängste von Sacherschließern, Ablehnung von komplexen maschinellen Verfahren und „Not-invented-here“-Haltungen. Jenseits dieser menschlichen Dimension ist zu berücksichtigen, dass Bibliotheken nicht in 100 oder 1000 Medientiteln denken, sondern oft in Hunderttausenden oder Millionen, das macht jede Kursänderung schwerfällig.

Erfreulicherweise haben die Verbünde und die Deutsche Nationalbibliothek 2006 Impulse gesetzt, die Richtung ist neu bestimmt, die erste Million digitale Inhaltsverzeichnisse im deutschsprachigen Bibliotheksraum prognostizieren wir für 2008. Dann fehlen nur noch 14 Millionen Sachtitel aus früheren Jahren, schätzen wir.

## **5 Best-Practice**

intelligentCAPTURE entwickelte sich zunächst dort, wo die Bibliotheksverbünde und zentrale Strukturen nicht oder nur schwach präsent waren. Innovationen kommen nicht aus den etablierten Zentren, sondern vom Rande her. Sie setzen sich aber erst durch, wenn die Zentren sie akzeptieren und fördern. Der GBV spielt hier für AGI die Rolle des Förderers und zieht mittlerweile daraus einen erheblichen

eigenen Nutzen. Staatliche Förderung von EU oder BMBF wurde bis heute verwehrt.



*Abbildung 3: intelligentCAPTURE mobile wird direkt zwischen engen Regalen eingesetzt. Der Scanner fährt zum Buch, die hohen Medientransportkosten werden stark minimiert. Über WLAN kommuniziert die Scanstation mit dem Bibliothekssystem, dandelon.com, dem Domino-Server und dem Internet. Die Mitarbeiter kommunizieren über Chat, IP-Telefonie, Application Sharing und eMail mit der Projektleitung. Vom Einsatz gibt es ein Video auf der AGI-Homepage.*

Heute ist intelligentCAPTURE ein sehr performantes und hochwertiges Verfahren zur Digitalisierung von Inhaltsverzeichnissen, deren maschineller Auswertung und über dandelon.com des Austausches mittlerweile über fünf Staaten. Es ist auch die erste mobile Version (siehe Abbildung 3) für Digitalisierung zwischen Regalen. Mit 42.000 Titeln gescannt, komplett verarbeitet und publiziert via GBV an drei Workstations in zwei Monaten an der SUB Hamburg (Mitte Februar bis Mitte April 2007) liegt die Messlatte für Qualität (Image, OCR, Indexierung) und Performance nicht gerade niedrig.

## **6      Entwicklungsperspektive**

dandelon.com integriert bereits National-, Landes- und Hochschulbibliotheken, Verlage und Buchhandel Staaten übergreifend. Weitere Vernetzungen, noch mehr Content, noch mehr Fokussierung, noch mehr Verfahren aus dem Information Retrieval werden die weitere Entwicklung kennzeichnen. Die Dokumentenstrukturanalyse und Informationsextraktion werden an Bedeutung weiter stark zunehmen. Geld- und Personalmangel bremsen noch immer das bereits Mögliche, doch wo ist es anders?

## **7      Literaturverzeichnis**

Hauer, Manfred: Vergleich der Retrievalleistungen von Bibliothekskatalogen gegen erweiterte und neue Konzept. In ABI-Technik Dez. 2005, S 295-301



# Einsatz automatischer Indexierungsverfahren in der Datenbank PSYINDEX\*

*Peter Weiland, Andreas Gerards & Michael Gerards*

Zentrum für Psychologische Information und Dokumentation (ZPID), Trier

## **Zusammenfassung**

Dieser Beitrag beschreibt die Implementierung, Funktionsweise und erste Ergebnisse einer Evaluation des automatischen Indexierungssystems AUTINDEX im Rahmen der Dokumentation psychologischer Literatur und Medien in der Datenbank PSYINDEX. Das System generiert auf Basis einer umfassenden Indikatorenliste für Thesaurusbegriffe aus den Abstracts und Titeln Deskriptorenvorschläge zur Unterstützung der intellektuellen inhaltlichen Erschließung von Dokumenten aus der Psychologie. Der Aufbau der Indikatorenliste sowie die technische und methodische Integration von AUTINDEX in den Dokumentationsablauf werden dargestellt. Im Anschluss werden kurz die Ergebnisse einer ersten Evaluation vorgestellt, bei der für 63 Dokumente die intellektuell gefunden Deskriptoren mit den automatisch generierten Deskriptorvorschlägen abgeglichen wurden. Ein kurzer Vergleich zwischen AUTINDEX und dem in den 80iger Jahren entwickelten System AIR/PHYS schließt den Beitrag ab.

## **I Einführung**

Im Zentrum für Psychologische Information und Dokumentation (ZPID) wird das Softwarepaket AUTINDEX (AUTomatic INDEXing) des IAI<sup>1</sup> zur automatischen Extraktion von Schlagworten aus deutschen und englischen Texten in der Unterstützung des Indexierungsprozesses eingesetzt: AUTINDEX generiert aus den Titeln, Abstracts und Autorenschlagworten eines Dokuments Deskriptorvorschläge, die dem Humanindexierer zur Auswahl angezeigt und von ihm – sofern er sie als zum Dokumentinhalt passend bewertet – übernommen werden. Das Projekt beinhaltete zum

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 413-422.

<sup>1</sup> Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung e.V.

einen die Entwicklung einer entsprechenden Schnittstelle zum Datenbanksystem STAR<sup>2</sup> und die Einbettung der automatischen Indexierung in den Erfassungsworkflow, zum anderen aber auch die Erweiterung bzw. Anreicherung des für die Verschlagwortung benutzten Thesaurus zur Verbesserung der Indexierungsergebnisse.

## **2 AUTINDEX**

### **2.1 Komponenten von AUTINDEX**

Die Architektur von AUTINDEX und seine Komponenten werden ausführlich in Ripplinger / Schmidt (2001) beschrieben. Die Software setzt natürlichsprachliche Analyseverfahren ein und besteht aus den folgenden Elementen

- linguistische Analyse MPro
- Evaluierung der Textelemente
- Ermittlung von Wortgruppen durch Oberflächenparsing
- Ergebnisausgabe

Im letzten Schritt Ergebnisausgabe wird das Analyseresultat mit den entsprechenden Thesaurusbegriffen in Verbindung gebracht und eine gewichtete Liste von Deskriptorvorschlägen ausgegeben.

Das Ziel einer Vor-Indexierung durch ein System wie AUTINDEX soll zum einen eine Zeitersparnis für den Humanindexierer sein, da dieser Vorschläge direkt übernehmen kann und nur in geringerem Umfang im Thesaurus suchen muss, zum anderen soll durch die Vorschläge eine konsistentere Indexierung zwischen den Humanindexierern erreicht werden.

### **2.2 Vorevaluationen**

AUTINDEX wurde im ZPID bereits 2004 erprobt. Diese erste Evaluation brachte im Vergleich zur intellektuellen Indexierung noch unbefriedigende Ergebnisse. Zum einen lag der Anteil an irrelevanten Deskriptoren zu hoch, zum anderen wurden von dem System wichtige Deskriptoren nicht vorgeschlagen, die sich nicht direkt aus dem Text ergeben, sondern nur durch eine Abstraktionsleistung des Indexierers erzeugt werden.

### **2.3 Maßnahmen zur Optimierung des Indexierungsergebnisses**

Die Vorevaluation hat gezeigt, dass der zugrunde liegende Thesaurus mit aktuell 5488 Deskriptoren in deutsch und englisch nicht ausreichend ist, damit AUTIN-

---

<sup>2</sup> Produkt der Firma Cuadra Associates, Inc.

DEX sinnvolle Vorschläge erzeugen kann. Daher wurde das kontrollierte Vokabular durch die Einführung so genannter Indikatoren erweitert. Diese Begriffe stehen in enger Beziehung zu den eigentlichen Deskriptoren, sind aber keine direkten Synonyme. Ein Beispiel aus Gerards et al. (2006) verdeutlicht dies:

Das englische APA-Thesaurus-Schlagwort *Acalculia* wird in der deutschen Version der PSYINDEX-Terms mit *Rechenschwäche* übersetzt. Als Synonym verweist der ebenfalls in den PSYINDEX-Terms aufgeführte Begriff *Rechenunfähigkeit* auf diesen Deskriptor. Ergänzend wurden nun zu diesen Begriffen folgende Indikatoren formuliert: *Akalkulie*, *Dyskalkulie*, *Dyskalkulia*, *Rechenstörung*, *mathematische Lernschwierigkeiten*, *rechenschwach*, *Probleme im Rechnen*, *Rechenprobleme*, *verzögerter Rechenerwerb*. Trifft AUTINDEX im Dokument auf einen Begriff, der in der Indikatorliste (bestehend aus dem Deskriptor, seinen Synonymen und den zusätzlichen Indikatoren) enthalten ist, wird – vorausgesetzt es werden bestimmte Gewichte und Schwellenwerte erreicht – der entsprechende Deskriptor vorgeschlagen.

Die Erzeugung der Indikatoren für die Thesaurusbegriffe war eine intellektuelle Aufgabe, auf der Grundlage von Fachwörterbüchern und auch vorhandener Dokumente in der Datenbank PSYINDEX. Insgesamt enthält der Thesaurus nun 23.661 Indikatoren<sup>3</sup>.

### 3 Einbettung der automatischen Indexierung in den PSYINDEX-Workflow

Das Ziel des Einsatzes der automatischen Indexierungssoftware AUTINDEX im ZPID ist die Unterstützung des Humanindexierers bei der inhaltlichen Erschließung von PSYINDEX-Datensätzen. Daher klinkt sich das System im Workflow direkt nach der formalen Erfassung ein, in deren Verlauf bibliographische Daten eingegeben bzw. importiert und überprüft werden und eventuell schon vorhandene Abstracts eingescannt werden.

#### 3.1 Workflow

Die Ermittlung der Deskriptorvorschläge mit AUTINDEX läuft im Batch-Betrieb, d.h. alle formal erfassten PSYINDEX-Dokumente, die die notwendigen Bedingungen für eine Verarbeitung erfüllen, werden nächtlich mit AUTINDEX verarbeitet und die Deskriptorvorschläge werden zu den entsprechenden Datensätzen hinzuge-

---

<sup>3</sup> Stand 11.01.2007.

fügt. Damit ein PSYINDEX-Datensatz verarbeitet werden kann, müssen mindestens ein Abstract (deutsch oder englisch) und der Titel jeweils mit Angabe der Sprache vorliegen. Von AUTINDEX bearbeitete Datensätze werden markiert und in späteren Durchläufen nicht berücksichtigt, es sei denn, dass der zuständige Auswerter dies explizit möchte, beispielsweise nach Veränderung des Abstracts.

### 3.2 Kommunikation zwischen AUTINDEX und der Datenbank PSYINDEX

Die Kommunikation zwischen AUTINDEX und der Datenbank PSYINDEX (Cuadra STAR) wird über die XML-Schnittstelle von STAR abgewickelt. Abbildung 1 zeigt den Ablauf schematisch.

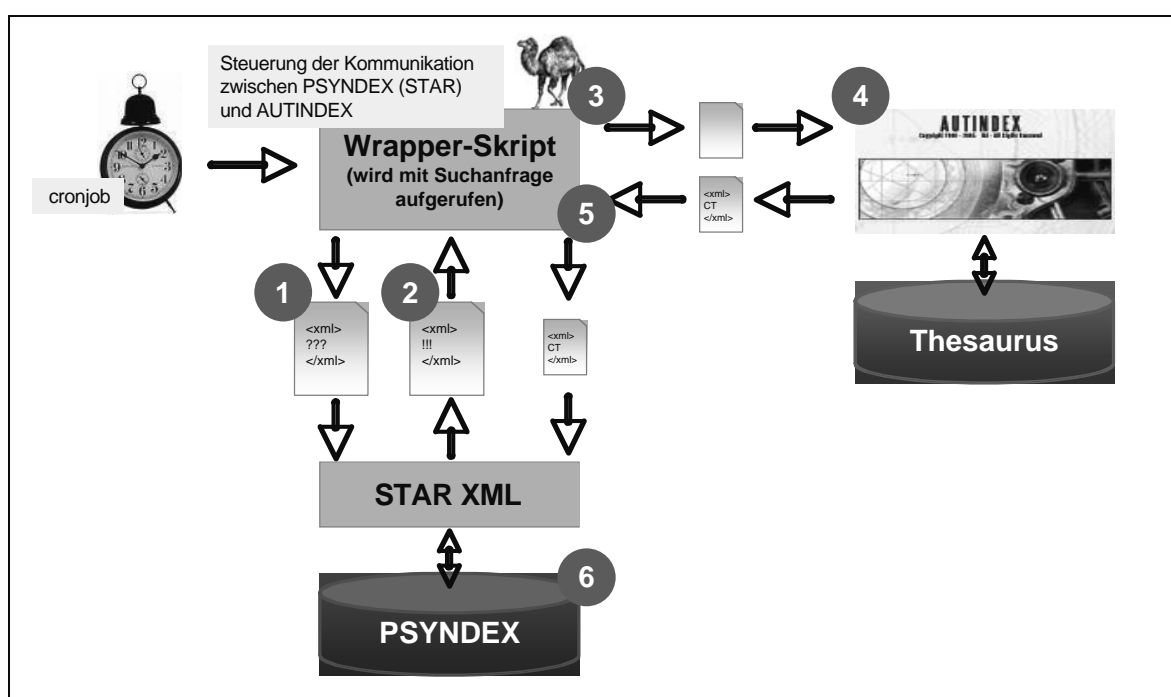


Abbildung 1: Kommunikation PSYINDEX – AUTINDEX

1. Ein durch einen cronjob angestoßenes Perl-Skript sendet eine XML-Anfrage mit einer Suche an STAR XML.
2. Als Antwort sendet STAR XML ein XML-Dokument, das alle Datensätze enthält, die den Bedingungen der Suchformulierung entsprechen<sup>4</sup>.
3. Das Perl-Skript verarbeitet das XML-Antwortdokument und splittet es für jeden Datensatz in eine einzelne Datei auf. Jede der Dateien wird anschließend an das AUTINDEX-Steuerungsskript übergeben.
4. AUTINDEX parst das übergebene XML-Dokument und ermittelt mithilfe des eingebundenen Thesaurus entsprechende Schlagworte. Als Ergebnis liefert AUTINDEX ein Dokument im STAR XML-Format zurück.

<sup>4</sup> Die Anzahl der zurück gelieferten Datensätze ist durch STAR XML auf 500 begrenzt.

5. Das „Wrapper“-Skript ergänzt das von AUTINDEX gelieferte Dokument um eine Markierung („Dokument ist bearbeitet“) und sendet dieses an STAR XML.
6. Das Dokument wurde in PSYINDEX um die automatisch erzeugten Schlagworte ergänzt.

### 3.3 Thesaurus

Die Basis für die Indexierung der PSYINDEX-Dokumente ist der um Deskriptoren erweiterte Thesaurus (siehe 3.2). Zur Nutzung des Thesaurus mit AUTINDEX muss dieser entsprechend kompiliert und eingebunden werden. Dieser Vorgang ist bei jeder Änderung/Erweiterung des Thesaurus notwendig.

### 3.4 Gewichte und Schwellenwerte

Über die Verteilung von Gewichtungen kann das Ranking der Deskriptorvorschläge beeinflusst werden. Neben der Differenzierung nach dem jeweiligen Feld (Titel/ Untertitel; Abstract/Nebenabstract; Autorenschlagworte) kann auch danach differenziert werden, welche Beziehung gefundener Begriff und Begriff im Text haben, z. B. Indikator, Deskriptor, Oberbegriff, Unterbegriff. Mit einem Schwellenwert wird festgelegt, ab welchem Gewicht ein Deskriptorvorschlag tatsächlich ausgegeben wird.

Sowohl für englische als auch für deutsche Dokumente werden die Beziehungen zwischen den Thesaurusbegriffen (*Broader Term*, *Narrower Term*, *Related Term*) für die Gewichtung momentan nicht berücksichtigt, da sich in Tests dadurch keine Qualitätsverbesserung ergeben hat.

### 3.5 Sicht des Indexierers auf vorindexierte Dokumente

Die Deskriptorvorschläge von AUTINDEX werden direkt in den PSYINDEX-Datensätzen gespeichert. Die Masken des Erfassungssystems wurden dahingehend erweitert, dass die Vorschläge einfach durch Markieren zu den *Controlled Terms* eines Dokuments hinzugefügt werden können. Abbildung 2 zeigt die Erfassungsmaske mit Deskriptorvorschlägen im STAR Client.

In der Weberfassungsmaske für PSYINDEX sind die Deskriptorenvorschläge der automatischen Indexierung zusätzlich mit einer Suche im Thesaurus verlinkt, so dass der Humanindexierer auch den Kontext eines Deskriptors sehen kann. In Abbildung 3 wird die „Umgebung“ des Deskriptors *Posttraumatic Stress Disorder* angezeigt.

Weiter		Abbrechen		Eintrag Einf.		Speichern		Entwurf	
Öffnen		Zurück		Inhaltliche Erfassung		Eintrag Löschen			
1 Abstract		2 Inhalt		3 Controlled Term		4 Titelübers.		5 Nebenabstr.	
6 Tests		7 Fehler							
186368: Hell, Benedikt, 2006, 58-78, Verwendung und Einschätzung von Verfahren der internen Personalauswahl und Personalentwicklung im 10 Jahres-Vergleich									
Controlled Term (engl oder germ)		Controlled Term (Übersetzung)		Gewichtet?		AUTINDEX-CT-Vorschläge			
Personnel Selection		Personalauslese		<input checked="" type="checkbox"/> gew.		Business Organizations Unternehmen (Wirtschaft)			
Employment Tests		Personalauslesetests		<input checked="" type="checkbox"/> gew.		Interviews Interviews			
Personnel Training		Personalschulung		<input checked="" type="checkbox"/> gew.		Personnel Selection Personalauslese			
Business Organizations		Unternehmen (Wirtschaft)		<input type="checkbox"/> gew.		Personnel Training Personalschulung			
Assessment Centers		Assessment-Center		<input type="checkbox"/> gew.		Personnel Evaluation Personalbeurteilung			
Job Applicant Interviews		Bewerbungsgespräche		<input type="checkbox"/> gew.					
Trends		Trends		<input type="checkbox"/> gew.					
Organizational Behavior		Organisationsverhalten		<input type="checkbox"/> gew.					
satzdeskript. deutsch									
				<input type="checkbox"/> gew.					
				<input type="checkbox"/> gew.					
phrase									
rends in use & evaluation of procedures for internal personnel selection & personnel development, 1993 vs 2003, increase in number of applied procedures & in frequency of use, greatest increase in assessment center structured interviews, 5 procedures for personnel									
<input checked="" type="radio"/> AUTINDEX durchführen bzw. wiederholen <input type="radio"/> nein <input type="radio"/> ja <input type="checkbox"/> Inhaltliche Erfassung unvollständig									
						Uncontrolled Terms (engl) Personnel Selection, Human Resource Development, Validity, Practicability, Acceptance, Follow-up Study Uncontrolled Terms (germ)  			

Abbildung 2: AUTINDEX Vorschläge in der STAR Client Maske

<b>Klassifikation und Deskriptoren</b>	
Klassifikation (SH)*	
3314   Interpersonal & Client Centered & Humanistic Therapy <span>Feld hinzufügen</span>	
Schlagwörter (CT)*	
Gestalt Therapy	<input checked="" type="checkbox"/> gew. Feld hinzufügen
Stuttering	<input type="checkbox"/> gew. Löschen
Posttraumatic Stress Disorder	<input type="checkbox"/> gew. Löschen
Family Relations	<input type="checkbox"/> gew. Löschen
AUTINDEX Schlagwörter	
Arguments	
Gestalt Therapy	
Posttraumatic Stress Disorder	
Stuttering	
Autorenschlagwörter englisch (UTG)	
keine Autorenschlagwörter vorhanden	
Autorenschlagwörter englisch (UTE)	
keine Autorenschlagwörter vorhanden	
Zusatzdeskriptoren (IT)	
	<input type="checkbox"/> gew. Feld hinzufügen
	<input type="checkbox"/> gew. Löschen
BT: Anxiety Disorders <b>Posttraumatic Stress Disorder</b> RT: Acute Stress Disorder RT: Adjustment Disorders RT: Combat Experience RT: Debriefing (Psychological) RT: Emotional Trauma RT: Stress Reactions RT: Traumatic Neurosis Deutscher Begriff: Posttraumatische Belastungsstörung Scope-Note:	

Abbildung 3: AUTINDEX-Vorschläge in PSYINDEX Weberfassung

## **4 Evaluation (November 2006)**

Im November 2006 wurde eine erste Evaluation von AUTINDEX durchgeführt – dabei wurde bei 63 Dokumenten ein Abgleich zwischen intellektuell vergebenen und automatisch generierten Deskriptoren vorgenommen. Grundannahme ist dabei, dass die vom menschlichen Indexierer vergebenen Deskriptoren den Dokumentinhalt in angemessener Weise beschreiben.

Die Ergebnisse dieser ersten Evaluation lassen sich folgendermaßen zusammenfassen (vergleiche Gerards et al., 2006):

- Im Schnitt kann ein Humanindexierer 3 Deskriptorvorschläge von AUTINDEX direkt übernehmen.
- Zusätzlich müssen 3-4 weitere Deskriptoren vergeben werden, die vom System nicht vorgeschlagen wurden.
- In der Mehrzahl der Dokumente schlägt AUTINDEX einen weiteren brauchbaren Deskriptor vor, der von einem menschlichen Indexierer nicht vergeben wurde.
- Das Indexat von AUTINDEX enthält im Durchschnitt einen weiteren Deskriptor, der – wenn auch nicht direkt verwendbar – in einer unmittelbaren Beziehung zu einem passenden Begriff steht.

## **5 Frühere Ansätze zur automatischen Indexierung in Fachinformationszentren**

### **5.1 AIR/PHYS**

Bereits Mitte der 80iger Jahre wurde das an der TH Darmstadt entwickelte Verfahren AIR/X als AIR/PHYS auf die Datenbank Physik des Fachinformationszentrums Karlsruhe angewendet. Indexiert wurden Datensätze mit englischsprachigen Titeln und Abstracts. Das System benutzte ein spezielles Lexikon, das Term-Deskriptor-Beziehungen für eine große Anzahl von Ein- oder Mehrworttermen des Anwendungsfeldes enthielt. Das Lexikon umfasste ungefähr 200.000 Ein- und Mehrwortterme, wovon 23.000 Deskriptoren sind. Zur Erzeugung des Lexikons wurden ca. 400.000 manuell indexierte Dokumente verarbeitet. Hierbei wurden zum einen die Beziehungen zwischen Deskriptoren und Wörterbucheinträgen erzeugt, zum anderen musste auch die Beziehung zwischen in der Physik gebräuchlichen Formeln und Deskriptoren hergestellt werden.

Zur Bestimmung eines Deskriptors mit AIR/PHYS für einen Datensatz werden folgende Schritte durchlaufen (Biebricher et al., 1988):

1. Textanalyse  
Zerlegung des Textes in Sätzen und einzelne Wörter; Rückführung auf Stammformen, Identifizierung von Stopp-Wörtern
2. Verarbeitung der Formeln: Ersetzung der Formeln durch die entsprechenden standardisierten Terme
3. Markierung der Terme, für die eine Beziehung zu einem Deskriptor existiert
4. Erstellung der Relevanzbeschreibungen  
Die Beschreibungen enthalten die Form des Terms und die Position, in der er im Text erscheint, die Art der Beziehung zwischen Term und Deskriptor und die aus den intellektuell vorindexierten Dokumenten berechnete Z-Relation zwischen Term und Deskriptor.
5. Berechnung des Gewichte
6. Korrektur der Gewichte durch Iteration der Schritte 4 und 5
7. Transformation des Ergebnisses  
In diesem Schritt werden die Deskriptoren dem Text zugewiesen, wenn ein bestimmter Schwellenwert erreicht wird.

Im Gegensatz zu dem für PSYINDEX eingesetzten AUTINDEX findet bei AIR/PHYS keine umfangreiche linguistische Analyse statt. AUTINDEX weist in der Analysephase jedem Wort im Dokument grammatikalische Informationen (z. B. Wortklasse) und semantische Merkmale zu. Darüber hinaus beherrscht das System auch eine Kompositaanalyse für die deutsche Sprache und die Erkennung von Mehrwortlexemen. AUTINDEX arbeitet sowohl mit deutsch- als auch mit englischsprachigen Dokumenten.

Eine Gemeinsamkeit beider Ansätze ist die Benutzung eines speziellen Lexikons, das im Falle von AIR/PHYS automatisch aus intellektuell vorindexierten Dokumenten erstellt wurde. Der Thesaurus von PSYINDEX, der mit 5488 Termen weniger umfangreich ist als der ca. 23.000 Begriffe umfassende Thesaurus für AIR/PHYS wurde hingegen intellektuell mit entsprechenden Indikatorbegriffen (siehe 3.2) erweitert. Im von AUTINDEX benutzten Lexikon gibt es keine Werte für die Beziehung zwischen Indikator und Thesaurusterm.

## **5.2 Evaluation von AUTINDEX in anderen Fachinformationseinrichtungen**

Die im ZPID eingesetzte Software wurde bereits in mehreren Projekten in Fachinformationseinrichtungen evaluiert:



- Im Rahmen des EU-Projektes BINDEX wurde AUTINDEX gemeinsam mit FIZ Technik und IEE/INSPEC in den Niederlanden von 2000 bis 2002 zur automatischen Indexierung zweisprachiger Texte weiterentwickelt und evaluiert. Eine ausführliche Beschreibung des Projektinhaltes und der Ergebnisse findet sich bei Nübel et al. (2002).
- Beim Hamburger Weltwirtschaftsarchiv (HWWA)<sup>5</sup> und der Zentralbibliothek für Wirtschaftswissenschaften (ZBW) wurde AUTINDEX in einem DFG-Projekt von September 2002 bis August 2004 zur automatischen Verarbeitung von Volltexten aus den Wirtschaftswissenschaften eingesetzt (siehe IAI, 2004).

## **6 Fazit und Ausblick**

Der Einsatz von automatisch generierten Deskriptoren kann den Indexierer bei seiner Arbeit unterstützen, insbesondere auch dann, wenn die Vorschläge des Systems Ausgangspunkt für das Finden weiterer Deskriptoren sind. Eine weitere Optimierung der Indexierungsleistung ist durch die Erweiterung bzw. auch Bereinigung der Indikatoren der einzelnen Thesaurusbegriffe zu erreichen. Darüber hinaus kann durch die Überarbeitung des englischen Thesaurus, für den bisher keine Indikatoren vorliegen, eine Verbesserung der Indexierungsqualität für englischsprachige Haupt- und Nebenabstracts erreicht werden.

Die bisher durchgeführten Evaluationen beschränkten sich auf eine recht kleine Anzahl von Dokumenten. Daher ist beabsichtigt, eine größere Anzahl von Dokumenten aus der Datenbank nachzuindexieren und das Ergebnis mit den intellektuell vergebenen Deskriptoren abzugleichen.

## **7 Literatur**

- Biebricher, P.; Fuhr, N.; Lustig, G.; Schwantner, M.; Knorz, G. (1988). The automatic indexing system AIR/PHYS – from research to applications. In Proceedings of the 11<sup>th</sup> Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Grenoble, France). Y. Chiaramella, Ed. SIGIR '88. ACM Press, New York, NY, 333-342. <http://doi.acm.org/10.1145/62437.62470>
- Gerards, M.; Gerards, A.; Weiland, P. (2006). Der Einsatz der automatischen Indexierungssoftware AUTINDEX im Zentrum für Psychologische Information und

---

<sup>5</sup> Das Hamburger Weltwirtschaftsarchiv wurde als Institut zum 31.12.2006 aufgelöst. Die Bibliothek wurde in die Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW) in Kiel integriert.

- Dokumentation (ZPID). Arbeitsbericht, online verfügbar unter <http://www.zpid.de/download/PSYINDEXmaterial/autindex.pdf> am 13.04.2007.
- IAI (2004). Abschlussbericht zum Projekt AUTINDEX (DFG-Geschäftszeichen: 554 922 (1) UV). Institut der Gesellschaft zur Förderung der angewandten Informationsforschung e.V, an der Universität des Saarlandes. Online verfügbar unter <http://www.iai.uni-sb.de/docs/AB-AUTINDEX.pdf> am 13.04.2007.
- Nübel, Rita; Pease, Catherine; Schmidt, Paul; Maas, Dieter (2002). Bilingual Indexing for Information Retrieval with AUTINDEX. In: LREC Proceedings, Las Palmas 2002. Online verfügbar unter <http://www.iai.uni-sb.de/~bindex/IrecNuebel.pdf> am 13.04.2007.
- Ripplinger, B. und Schmidt, P. (2001). AUTINDEX: an automatic multilingual indexing system. In: Proceedings of the 24<sup>th</sup> Annual international ACM SIGIR Conference on Research and Development in information Retrieval (New Orleans, Louisiana, United States). SIGIR '01. ACM Press, New York, NY, 452.  
<http://doi.acm.org/10.1145/383952.384093>

# **Zwischenbericht aus der Virtuellen Fachbibliothek Sportwissenschaft\***

*Siw Waffenschmidt*

Projektkoordinatorin der Virtuellen Fachbibliothek Sportwissenschaft,  
Zentralbibliothek der Sportwissenschaften der Deutschen Sporthochschule Köln

## **I Einleitung**

Die Virtuelle Fachbibliothek Sportwissenschaft (ViFa Sport) dient der Ausweitung und Verbesserung der überregionalen Literaturversorgung mit konventionellen und digitalen Informationsressourcen.

Potentielle Kundinnen und Kunden sind alle Personen, die ein Interesse an sportwissenschaftlichen Informationen haben. Dies umfasst zum einen die Wissenschaft, Forschung und Bildung in der Sportwissenschaft. Zum anderen aber auch Personen aus dem organisierten Sport wie Sportler, Übungsleiter und Trainer sowie sportinteressierte Personen.

Der Startschuss für den Aufbau der ViFa Sport fiel im Januar 2006 bei einem Treffen der Projektpartner in Köln, zu denen die Zentralbibliothek der Sportwissenschaften (ZBS) als Projektleitung, das Bundesinstitut für Sportwissenschaft (BISp), das Institut für Angewandte Trainingswissenschaft (IAT), die Deutsche Vereinigung für Sportwissenschaft (dvs) und die Bibliothek der Friedrich-Ebert-Stiftung (FES) gehören. Grundlegende Ziele für den Aufbau der ViFa Sport sind:

- die Zusammenführung der verschiedenen sportwissenschaftlichen Informationsbestände in einem frei zugänglichen Internetauftritt,
- eine gemeinsame Recherche über diese Bestände,
- der direkte Zugriff auf die Bestände ohne Medienbruch (Verlinkung von digitalen Volltexten, Online-Fernleihe, elektronische Dokumentlieferung),
- die Nachhaltigkeit des Angebots sowie die Integration in das Internetportal vascoda.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 423-426.

Folgende Informationsressourcen stellen den sportwissenschaftlichen Kern des Angebots dar:

- der zentrale SSG-OPAC der ZBS sowie weitere relevante Bibliothekskataloge (Auszüge aus den Katalogen der Bibliothek der FES und der Deutschen Zentralbibliothek für Medizin [ZBMed]),
- medienübergreifende Fachdatenbanken: das Angebot SPORTIF (bereits über vascoda vorhanden) mit den Datenbanken SPOLIT, SPOMEDIA und SPOFOR des BISp; SPOWIS und SPONET des IAT, Ausschnitt aus dem Fachinformationssystem Bildung (FIS Bildung) des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF),
- der wissenschaftlich geprüfte Fachinformationsführer Sport als Quality Controlled Subject Gateway zu sportwissenschaftlichen Internetressourcen (über das Portal SPORTIF des BISp im Aufbau),
- der Current-Contents-Dienst, der an der ZBS erstellt und den sportwissenschaftlichen Teil der „Online Contents Sondersammelgebiete“ (OLC-SSG) des Gemeinsamen Bibliotheksverbundes (GBV) bilden wird,
- Online-Volltexte (Retrodigitalisate und E-Publikationen): im Projektverlauf digitalisiert werden die Zeitschrift „Theorie und Praxis Leistungssport“ (vom IAT), die dvs-Schriftenreihe sowie der FES-Sonderbestand Arbeitersport,

Die erste Hälfte des Projektes war geprägt von der Diskussion über die einzusetzende Suchtechnologie. Nach intensiver Prüfung wurde der Einsatz der grundsätzlich präferierten FAST-Suchmaschinenteknologie aufgrund des engen Zeitfensters auch in Hinblick auf zu klärende rechtliche Rahmenbedingungen zurückgestellt und die Realisierung mit einer Metasuche über die Information Portal Suite (IPS) beschlossen. Das Hosting der ViFa Sport wird durch das Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz) gewährleistet, die Hardware und Software für die Metasuche bereitstellen.

Dieser Halbzeitbericht präsentiert ausgewählte Arbeitsbereiche aus der Vifa Sport wie die Terminologiarbeit, den Aufbau eines Current-Contents-Dienstes sowie die Schaffung von nachhaltigen Strukturen.

## **2 Terminologiarbeit**

Zur Verbesserung der Suchergebnisse wurden die einzelnen Instrumente der inhaltlichen Erschließung (Schlagwortliste des BISp, Thesaurus des IAT, Systematik der

ZBS und Sachschlagwortliste der FES) per Konkordanzerstellung miteinander verknüpft.

Die Terminologie-Arbeit wurde in eine eigens dafür entwickelte Datenbank umgesetzt, in der jede Institution pro Datensatz ein eigenes Deskriptorenfeld hat. Die Konkordanzerstellung stellt einen indirekten Termtransfer dar; es wurde also eine Verbindung zwischen den Dokumentationssprachen hergestellt. Die Konkordanz-erstellung ist eine Äquivalenz- und Ähnlichkeitsrelation, die zusätzlich auch logische UND-Verknüpfungen der Deskriptoren (Beispiel: aerob UND Training) beinhaltet. Zusätzlich wurde die Terminologie-Datenbank mit folgenden Ergänzungen angereichert: Synonyme, Rechtschreibvarianten, grammatische Formen, Abkürzungsaufösungen, Bindestrichauflösungen, englische Übersetzungen.

### **3 Current-Contents-Dienst**

In Absprache mit dem Gemeinsamen Bibliotheksverbund (GBV) in Göttingen wurden in der ZBS bislang 46.000 Artikeldaten durch Bearbeitung der Inhaltsverzeichnisse von über 60 Zeitschriften gewonnen.

Für den Current Contents-Dienst wurden Zeitschriftenaufsatzdaten aus ausgewählter Zeitschriften gesammelt.

Zusätzlich zu den Artikeldaten, die in der ZBS erfasst wurden, können aus dem GBV ca. 270.000 Artikeldaten zur Verfügung gestellt werden. Dabei handelt es sich um die Zeitschriften, welche die ZBS abonniert hat und deren Inhalte bereits im GBV als Digitalisate vorlagen.

Die durch DFG-Mittel gewonnenen Artikeldaten stehen zusammen mit den GBV-Artikeldaten in einer sportbezogenen Datenbank zur Verfügung.

### **4 Nachhaltigkeit**

Die Schaffung nachhaltiger Strukturen sind weitere Ziele der ViFa Sport. Es wurde daher in der Projektgemeinschaft beschlossen, für die Zeit nach der Projektförderung durch die DFG eine Kooperationsvereinbarung zur Weiterführung der ViFa Sport zu erstellen. Dort werden zukünftige Ziele, Aufgaben, Zuständigkeiten, Kosten, Verantwortlichkeiten, u. ä. schriftlich dokumentiert.

## **5 Zusammenfassung**

Die Erfahrung aus zwölf Monaten Projektlaufzeit zeigt, dass das Projekt ViFa Sport eine Herausforderung für alle Projektpartner der Antragsgemeinschaft ist. Die Eröffnungsveranstaltung der ViFa Sport erfolgt anlässlich des dvs-Hochschultags am 26. – 28. September 2007 in Hamburg.

### **Kontakt**

Siw Waffenschmidt – Projektkoordinatorin "ViFa Sport" -  
Zentralbibliothek der Sportwissenschaften  
der Deutschen Sporthochschule Köln  
Carl-Diem-Weg 6 , 50933 Köln  
Tel.: 0221/4982-3390 ; Fax: 0221/4982-8070 ;  
mailto: *waffenschmidt@dshs-koeln.de*  
*<http://www.vifasport.de/>*

# Information folgt Prozess.

## Prozessorientierte Informationssysteme zur Qualitätssicherung kleinerer und mittlerer Forschungseinrichtungen\*

*Jutta Marx*

Qualitätsbüro Dr. Marx  
Walter-Kollo-Str. 14, 14513 Teltow  
Tel.: 03328-337397, Fax: 03328-338725  
*marx@qb-marx.de*

### **Zusammenfassung**

Am Beispiel des Wissensmanagementsystems knowWare wird das Prinzip prozessorientierter Informationssysteme vorgestellt. Diese kombinieren die Visualisierung von Geschäfts- und Arbeitsprozessen mit Dokumenten- und Wissensmanagement-funktionalitäten. Sie eignen sich daher sehr gut zur dezentralen Erfassung wissenschaftlicher Leistungskennziffern, wie sie von allen wissenschaftlichen Forschungs- und Serviceeinrichtungen im Zuge der Forschungsevaluation zu dokumentieren sind. Gerade kleinere und mittlere Einrichtungen, bei denen dezentrale Workflows gut zu etablieren sind, können durch den Einsatz prozessorientierter Informationssysteme entscheidend bei der Evaluationsvorbereitung und Qualitätssicherung profitieren.

Informationen über den Forschungs- und Wissenschaftsbetrieb, seine Ausstattung, seine Produkte und seine Effektivität sind in den letzten Jahren zunehmend in den Fokus von Wissenschaft, Politik und Öffentlichkeit gerückt. Auslöser dafür war weniger das Interesse der Betroffenen selbst als vielmehr äußerer Druck in Form von Evaluationen und Rankings.

Forschungsevaluation ist momentan eines der Topthemen. Stichworte wie Hochschulranking, Bologna-Prozess und Exzellenzinitiative bestimmen die Wissenschaftsrubriken aller Medien. Doch nicht nur Universitäten und Hochschulen sind

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 427-431.

von dieser Entwicklung betroffen, auch außeruniversitäre Forschungs- und Serviceeinrichtungen sind Gegenstand der allgegenwärtigen Effizienzdebatte.

Begegnet wird diesen Herausforderungen mit diversen Konzepten zur Qualitätssteigerung und –sicherung wie bspw. Programmbudgets, Kosten-Leistungs-Rechnung oder Zielvereinbarungen<sup>1</sup>.

Darüber hinaus werden allerorten fieberhaft wissenschaftliche Leistungskennziffern erfasst, um die eigene Leistungsfähigkeit und Exzellenz zu demonstrieren. Dieses Unterfangen ist jedoch nicht ganz trivial.

Neben vielen anderen Problemen wie z. B. der Aussagekraft bestimmter Kennziffern und Indikatoren für die einzelnen Wissenschaftsdisziplinen (s. Hornbostel 2005) liegt eine Hauptschwierigkeit in der mangelnden Motivation der „Kennzahlenerzeuger“, d.h. der einzelnen Wissenschaftler/innen. Zumindest wenn der enge Bereich der wissenschaftlichen Publikationen verlassen wird, bedeutet die Leistungskennziffernerfassung für sie eine zusätzliche Arbeitsbelastung, die in der Realität oft nur widerwillig, unregelmäßig und mit mangelnder Sorgfalt geleistet wird.

Im Umfeld der Hochschulen und Universitäten haben sich inzwischen weitere Einrichtungen (wie z. B. die Hochschul-Informationssystem GmbH HIS, [www.his.de](http://www.his.de) oder das Institut für Forschungsinformation und Qualitätssicherung IFQ, [www.forschungsinfo.de](http://www.forschungsinfo.de)) gegründet, die sich des Themas „Forschungsinformation“ annehmen. Die Softwarelösungen, die auf dem Hochschulbereich Anwendung finden, sind für kleinere und mittlere Einrichtungen aber meist völlig überdimensioniert und entsprechen nicht deren spezifischen Bedürfnissen.

Allerdings haben diese Einrichtungen gegenüber den Hochschulen den Vorteil der größeren Homogenität. Die Gruppe der Wissenschaftler/innen ist überschaubar, die Organisationseinheiten sind weniger autark und daher leichter steuerbar. Eine dezentrale Eingabe oder Zulieferung der erforderlichen Informationen kann dauerhaft zu einer qualitätsgesicherten Informationsbasis führen, wenn es gelingt, stabile Workflows zu etablieren. Unabdingbar sind Mehrwertanreize, um die nachhaltige Akzeptanz der Maßnahmen bei den meist nur schwach strukturiert agierenden Wissenschaftlern/innen zu unterstützen (Walter 2005).

Nach der exakten Definition der zu erhebenden Kennzahlen inklusive einer genauen Beschreibung ihrer Parameter (z. B. die Differenzierung wissenschaftlicher Pub-

---

<sup>1</sup> Eine umfassende Darstellung der von den deutschen Hochschulen eingesetzten Verfahren zur Qualitätssicherung bietet das Netzwerk Qualitätssicherung und Qualitätsentwicklung an Deutschen Hochschulen evaNet (<http://evanet.his.de>).



likationen und ihrer Ansetzungsformen, s. z. B. Marx 2002) empfiehlt es sich, nach den Regeln des prozessorientierten Qualitätsmanagements (s. DGQ 2005) Prozesse zur Erhebung der Kennziffern zu entwickeln.

Abbildung 1 zeigt einen solchen Prozess, der beschreibt, wie mit Hilfe der Literaturverwaltungssoftware EndNote des Anbieters Adept Scientific ([www.adeptscience.com](http://www.adeptscience.com)) eine Untergruppe wissenschaftlicher Leistungskennziffern (Publikationen, Vorträge, Veranstaltungsbesuche, Graduierungen) durch den einzelnen Wissenschaftler zu erfassen ist.

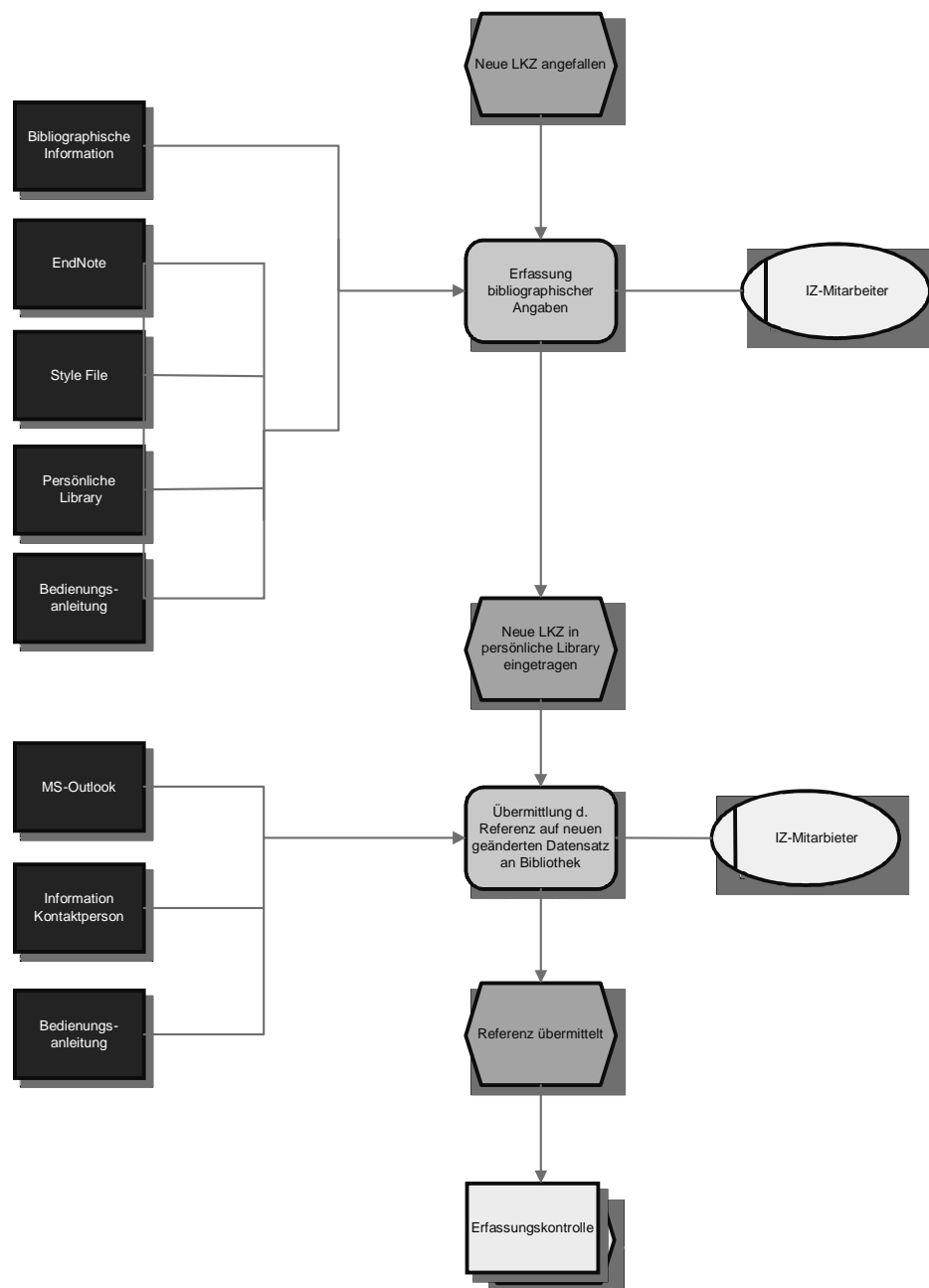


Abbildung 1: Basisprozess zur Erfassung wissenschaftlicher Leistungskennziffern am Beispiel des Informationszentrums Sozialwissenschaften Bonn

Zur korrekten Erfassung sind nun eine Menge unterschiedlicher Informationen, wie z. B. Vorgaben zu Ansetzungsformen, Erfassungsrichtlinien, Beispiel-Datensätze, Ansprechpartner u. ä. erforderlich, die in anderen Dokumenten und u. U. in anderer Form (z. B. Papiausdruck) vorliegen (in Abb. 1 die in Rechtecken bezeichneten Informationsobjekte).

Eine Ablage von Informationsobjekten in einer prozessorientierten Form hilft hier, eine natürliche Ordnung herzustellen und nachhaltig zu gewährleisten. Eine Software, die dies unterstützt ist z. B. das Modul eQs des Wissensmanagementsystems knowWare ([www.processware.de](http://www.processware.de)). Die webbasierte Intranetlösung eQs visualisiert die Geschäfts- und Arbeitsprozesse und zeigt zu jeder einzelnen Aktivität die relevanten Erklärungen, Dokumente und verantwortlichen Personen an. Das Modul unterscheidet einen Arbeitsbereich zum Austausch von geänderten Dokumenten und einen qualitätsgesicherten Ergebnisbereich, in dem nur durch festgelegte FreigabeprozEDUREN Informationen bereitgestellt werden können.

Abbildung 2 zeigt eine Prozessdarstellung aus dem Anwendungsbereich Psychiatrie des Landschaftsverbandes Westfalen-Lippe.

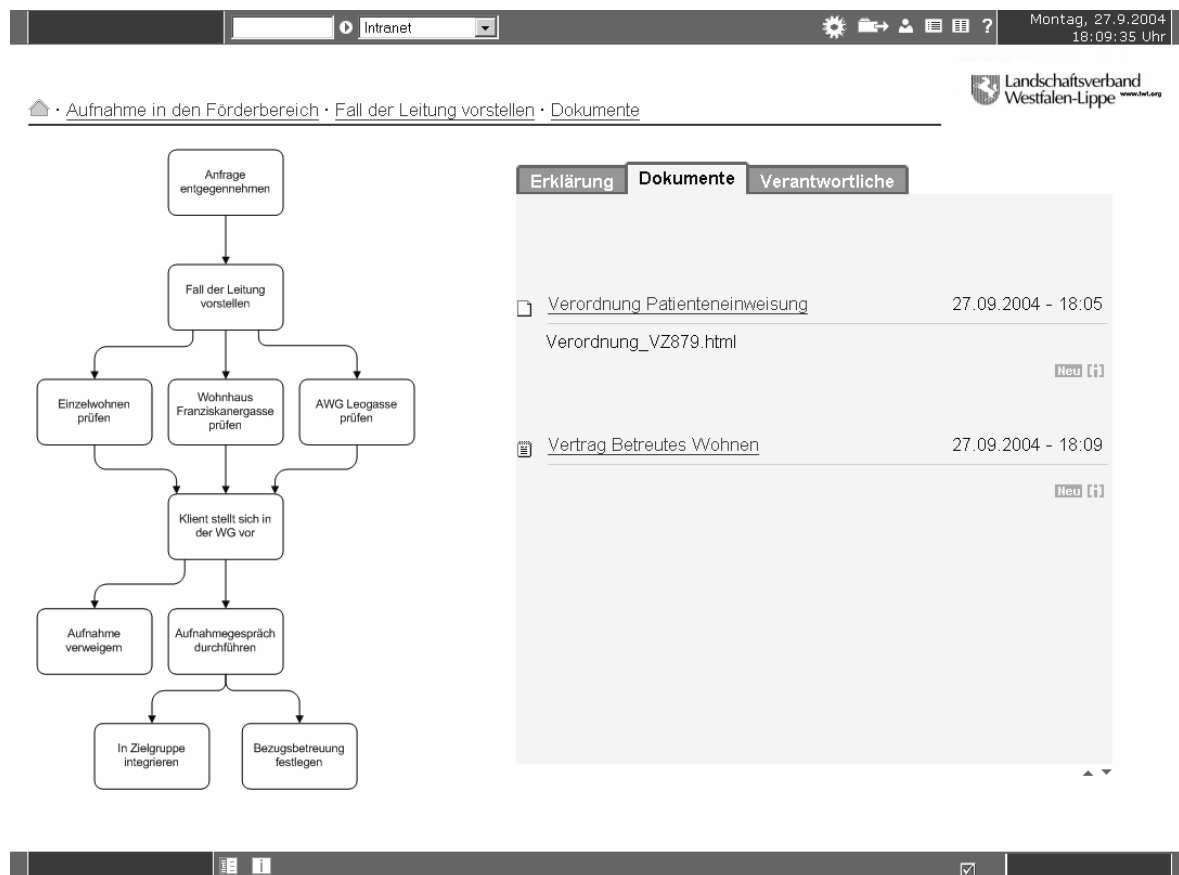


Abbildung 2: knowWare-Modul „eQs“ am Beispiel des Landschaftsverbandes Westfalen-Lippe

Die korrekte Abarbeitung einer Aktivität wird in der gegenwärtigen Version noch nicht kontrolliert. Die Prozessdarstellung dient ausschließlich der Visualisierung und der Zuordnung von Informationen und Dokumenten zu den einzelnen Prozessschritten.

Prozessorientierte Informationssysteme in der beschriebenen Form existieren bislang für die Bereiche Medizin, Psychiatrie und Jura. Es wird momentan geprüft, ob eine Umsetzung für die Erfassung und Bereitstellung von Forschungsinformationen konzeptuell und praktisch machbar ist. Dazu gehört auch die direkte Edition von Dokumenten innerhalb des Systems ohne explizites Upload ebenso wie der Aufruf externer Anwendungen.

Gelingt dies, so können prozessorientierte Informationssysteme gerade kleinere und mittlere Forschungseinrichtungen entscheidend bei der Evaluationsvorbereitung und beim Qualitätsmanagement unterstützen.

## Literatur

- Deutsche Gesellschaft für Qualität e.V. (2005): Prozessmanagement für Praktiker – Leitfaden für das Erkennen, Beschreiben, Bewerten, Umsetzen und Verbessern von Prozessen. DGQ-Band 14-26. Berlin.
- Hornbostel, S. 2005: Kennzahlen als Informations- und Steuerungsinstrumente – ein Methodenvergleich. S. 25-34 in: AG Fortbildung im Sprecherkreis der Universitätskanzler (Hg.): Einsatzmöglichkeiten und Grenzen von Kennzahlen als Informations- und Steuerungsinstrumente. Materialien Nr. 94. Weimar, Bauhaus-Universität: Universitätsverlag.
- Marx, J. (2002): Konzept zur Erfassung der Forschungs- und Serviceaktivitäten des Informationszentrums Sozialwissenschaften (Gruppe III: Wissenschaftliche Leistungen). IZ-Arbeitsmaterial Nr. 14. Bonn.
- Walter, R. (2005): Kommunikationsbeschleuniger in der virtuellen Wissenschaft. Studie im Auftrag des Bundesministeriums für Bildung und Forschung (zu beziehen über TIB Hannover unter <http://opc4.tib.uni-hannover.de>).



# **MathematikerInnen machen sich nützlich: Projektinformationen und eine Expertendatenbank der angewandten Mathematik\***

*Robert Roggenbuck*

Konrad Zuse Zentrum für Informationstechnik Berlin (ZIB)

## **Zusammenfassung**

Es steckt im Alltag mehr Mathematik als im Allgemeinen sichtbar ist. Nicht nur bei Banken, Versicherungen und Ingenieuren spielt Mathematik eine erhebliche Rolle, sondern auch in vielen anderen Bereichen. Im Bereich der angewandten Mathematik wird deutlich wie sehr Mathematik unseren Alltag durchdringt. Doch wer weiß schon, was MathematikerInnen im Bereich der angewandten Mathematik leisten? Und noch weniger kennen den Begriff „angewandte Mathematik“. Eine wichtige Frage für die mathematischen Experten (und auch für andere Wissenschaften) stellt sich daher bezüglich der Wahrnehmung in der Öffentlichkeit (und damit in gewisser Weise auch der Existenzberechtigung) – besonders in den Bereichen Industrie, Dienstleistungen und Förderinstitutionen. In diesem Beitrag wird gezeigt wie zwei wichtige Informationslücken geschlossen werden: Informationen über Forschungsprojekte und die Informationsquelle Mensch.

## **I Einleitung: Mathematik in der Öffentlichkeit**

„Mathematik ist überall“ – so titelte 2005 der Mathematiker Norbert Herrmann, und trug mit seiner Veröffentlichung zur Popularisierung der Mathematik bei (Herrmann 2005). Nicht nur in den Bereichen, die als „zahlennah“ bekannt sind, spielt Mathematik eine erhebliche Rolle, sondern auch bei der Ausrichtung von Mobilfunkantennen (vgl. Eisenblätter et al. 2002), der Fahrplangestaltung von Bussen (vgl. Löbel 1999), der Erstellung von Bildern aus Tomographiedaten (vgl. Gardner et al. 1999), oder Entwicklung von bodengestützten Messverfahren zur Bestimmung von FCKW-Anteilen (vgl. Böckmann et al. 2001), um nur einige we-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 433-444.

nige Beispiele zu nennen<sup>1</sup>. Die Spannweite der Themen ist noch weiter als diese Beispiele zeigen. Die angewandte Mathematik kann sehr gut verdeutlichen, wie sehr unser Alltag von Mathematik durchdrungen ist. Doch die Leistungen der MathematikerInnen im Bereich der angewandten Mathematik bleiben üblicherweise verborgen. Und noch weniger kennen den Begriff „angewandte Mathematik“.

Angewandte Mathematik ist heute im wesentlichen Forschung auf Projektbasis, und wird, sofern sie öffentlich gefördert wird, häufig gemeinsam mit Partnern aus Industrie oder dem Dienstleistungsbereich durchgeführt (z. B. im Rahmen des BMBF-Mathematikprogramms seit 1993<sup>2</sup> oder auch der DFG<sup>3</sup>).

Doch wie erfährt die Öffentlichkeit von den Projektergebnissen und wie kann sie davon profitieren? Erste Ansätze waren Publikationen in Buchform – mit deutlichen Nachteilen: Einschränkungen hinsichtlich des Umfangs und der Zielgruppe, Veröffentlichung mit starkem Zeitverzug. Als besser und zeitgemäßer bot sich das Internet als Publikationsform an.

Genau aus dieser Erkenntnis heraus entstand im Jahre 2001 das Projekt Math&Industry<sup>4</sup>, das die Aufgabe hatte, umfassend über das BMBF-Mathematikprogramm im Internet zu informieren – mit dem Hintergedanken einen Informationsdienst für die „mathematische Gemeinde“ zu schaffen, aber auch um für Mathematik und ihre Expertise für Industrie und Dienstleistungen zu werben.

## 2 Projektpräsentation mit Math&Industry

Am Beginn eines jeden Vorhabens steht die Konzeption. Verschiedenen Anforderungen galt es gerecht zu werden:

- Es soll eine Lösung gefunden werden, die langfristig mit geringem Betreuungsaufwand auskommt. Das Konzept sieht vor, dass die Projekte eigenverantwortlich ihre Informationen bereitstellen. Dazu müssen die Projekte gewonnen werden.
- Die Informationen sollen für unterschiedliche Zielgruppen aufbereitet werden:

---

<sup>1</sup> Einen umfassenden Einblick in die Anwendungsbereiche von Mathematik bietet das Portal von Math&Industry: <http://www.mathematik-21.de/>.

<sup>2</sup> Mathematik für Innovationen in Industrie und Dienstleistungen, <http://www.fz-juelich.de/ptj/mathematik>

<sup>3</sup> Matheon - Mathematik für Schlüsseltechnologien: Modellierung, Simulation und Optimierung realer Prozesse, <http://www.matheon.de/>

<sup>4</sup> Zur Zeit bearbeitet vom ZIB unter Mitarbeit des Instituts für wissenschaftliche Information (IWI) in Osnabrück und gefördert durch das BMBF.

- MathematikerInnen
- andere WissenschaftlerInnen und Fachleute (z. B. Ingenieure) aus der Industrie und dem Dienstleistungsbereich
- die fachlich interessierte Öffentlichkeit (Studenten, Lehrer, Fachjournalisten, ...)
- Manager und andere Entscheidungsträger aus den Anwendungsgebieten
- PolitikerInnen und Förderorganisationen (denn die Projekte wurden mit Steuermitteln finanziert)
- Durch die unterschiedlichen Zielgruppen ergibt sich die Anforderung nach unterschiedlichen Detailtiefen der Darstellung.
- Der (zusätzliche) Erstellungsaufwand der Internetpräsentationen soll minimal sein.
- Die Informationen (= Problemlösungen) müssen im Internet einfach und schnell gefunden werden können.

Als grundlegendes Element der Lösung wurde ein dezentraler Ansatz gewählt. Das Konzept besteht aus einem Portal als zentrale Informationsquelle, die über vielfältige Aufbereitungen der Projektinformationen zu den Projekten führt.

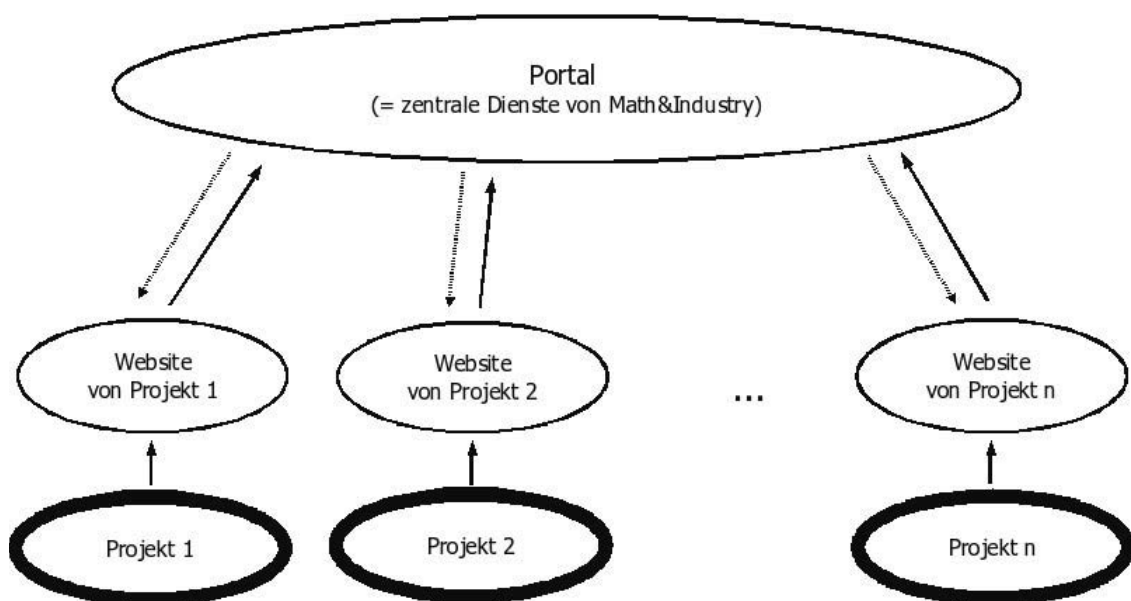


Abbildung 1: Dezentrale Informationsstruktur

Dezentral heißt hier vor allem, dass die Bereitstellung der Informationen (einschließlich der Websiteerstellung) von den Projekten geleistet wird und das Portal für die Bündelung und weitere Aufbereitung der Informationen zuständig ist, einschließlich der Verweise auf die Websites der Projekte (siehe Abbildung 1).

Somit ist die Lösung in die Bereiche Projektdarstellung und Portal geteilt. Die Teile tragen wie folgt zur Lösung bei.

## 2.1 Die Projektwebsites

Das Schlüsselwort für die Projektwebsites heißt „Strukturierung“. Durch eine einheitliche Gliederung der Informationen (Klassifizierung) wird die Aufbereitung der Projekte für die einzelnen Zielgruppen ermöglicht (grobe inhaltliche Strukturierung) und projektübergreifend den Besuchern der Websites eine einfache Orientierung geboten (Strukturierung des Erscheinungsbildes). Für die Informationen einer Klasse, z. B. von Publikationen, wurden spezielle Erschließungsmethoden (Metadatenschemata) definiert (feine inhaltliche Strukturierung). Diese können dann für die automatische Auswertung der Informationen und den Aufbau spezieller Informationsdienste genutzt werden.

Die Metadatenprofile, auch „Application Profiles“ genannt, legen fest, wie Aussagen über ein Objekt zu formulieren sind. Metadatenprofile wurden z. B. für die Klassen „Publikationen“, an den Projekten beteiligte „Mitarbeiter“ und „Firmen“, „Software“, „Veranstaltungen“, „Glossareinträge“ und „Verweise“ erstellt. Im Detail bedeutet das, festzulegen

- durch welche Informationen ein Objekt (wie z. B. eine Publikation) beschrieben wird,
- mit welchem Metadatenvokabular diese Informationen abgebildet werden (bevorzugt werden dabei existierende Standards<sup>5</sup> verwendet; falls sich im Einzelfall keine geeignete Vokabel findet, muss eine eigene im Rahmen eines eigenen Schemas entwickelt werden<sup>6</sup>)
- und auch die Relation der einzelnen Information zueinander (wie z. B. die Zuordnungen, dass ein Autor eine Emailadresse hat, eine Publikation mehrere Autoren haben kann, aber eine Publikation selbstverständlich keine Emailadresse hat).

Diese Profile bilden den Rahmen für die Strukturierung der Informationen und sind die Grundlage zur Errichtung intelligenter Dienste.

Die Metadatenprofile wurden in RDF/XML entwickelt. Die Implementierung der Metadatenschemata für konkrete Objekte ist ohne zusätzliche Werkzeuge aufwän-

---

<sup>5</sup> Weit verbreitete standardisierte Vokabulare sind zum Beispiel von der „Dublin Core Metadata Initiative“ (<http://dublincore.org/>) entwickelt worden und auch die „Web Ontology Language“ (OWL: <http://www.w3.org/TR/owl-ref/>).

<sup>6</sup> Als Beispiel für Eigenentwicklungen siehe die Vokabulare, die vom IWI entwickelt wurden: <http://www.iwi-iuk.org/material/RDF/Schema/>



dig, fehleranfällig und erfordert zudem einige Kenntnisse über die Extensible Markup Language (XML) und das Resource Description Framework (RDF)<sup>7</sup>, zwei Web-Standards, mit denen eine flexible Beschreibung und inhaltliche Erschließung der Objekte vorgenommen werden kann.

Um den Projekten die umfangreiche Arbeit der Erstellung einer solchen Webpräsentation zu erleichtern bzw. überhaupt erst zu ermöglichen, wurde ein spezialisiertes und leicht zu bedienendes Content Management System, namens WebSiteMaker, entwickelt<sup>8</sup>. Der WebSiteMaker steht online zur Verfügung, so dass die Software von den Projekten nicht installiert werden muss.

## 2.2 Das Portal

Das zentrale Element des Portals ist das Konzept des „Dienstes“. Als Dienste werden hier alle Möglichkeiten verstanden, die die Informationen der Projekte zusammenfassen und in neuer Form aufbereiten bzw. auf die Projekte verweisen. Grundlage der Dienste sind die (Meta-) Daten der Projektwebsites. Als Beispiel für Dienste seien hier schlichte Projektlisten, eine Volltextsuche über Portal und Projektwebsites, ein Glossar<sup>9</sup>, oder eine Publikations- sowie eine Softwaredatenbank genannt. Außerdem wird über das Portal dafür gesorgt, dass die Projekte über Google und andere Suchmaschinen gefunden werden. Und natürlich finden sich hier Informationen über das Projekt Math&Industry selbst.

Eine Besonderheit in der Implementierung dieser Informationsstruktur ist die Verwendung von Methoden des Semantic Web. Die gesamten Projektinformationen sind als Metadaten in RDF beschrieben und als XML-Dateien kodiert. Diese Beschreibungen sind mit den jeweiligen Webseiten verknüpft – XHTML- und RDF-Dateien verweisen aufeinander. Auf diese Weise können die RDF-Dateien durch einen Webcrawler eingesammelt werden, um sie auf dem Server des Portals auszuwerten. Diese Auswertung geschieht zurzeit jedoch noch nicht anhand der RDF-Tripel mit denen beliebige Aussagen der Form „Subjekt – Prädikat – Objekt“ modelliert werden können, sondern durch das Parsen der RDF-Dateien mit Hilfe der Erstellungswerkzeuge (der MIPMs<sup>10</sup>; sie sind Komponenten des WebSiteMakers).

---

<sup>7</sup> <http://www.w3.org/RDF/>

<sup>8</sup> Näheres zu diesem Aspekt von Math&Industry vgl. (Roggenbuck et al. 2006).

<sup>9</sup> Über die spezielle Rolle von Glossaren bei Math&Industry siehe (Roggenbuck 2005).

<sup>10</sup> Weitere Informationen zu den „Math&Industry Presentation Makern“ (MIPM) finden sich unter <http://www.mathematik-21.de/software/software.shtml#websiteerstellungstools>; dort stehen sie auch zum Download zur Verfügung.

## 2.3 Suchen und Finden

Findet die Öffentlichkeit nun, nachdem dies realisiert ist, die Antworten auf ihre Fragen und vor allem die Lösungen ihrer Probleme? Grundsätzlich schon, denn das Portal verweist auf Projekte und über die Projekte finden sich darüber hinaus Institutionen und Personen, die noch mehr wissen als auf der Projektwebsite steht. Wie gut das gelingt, hängt einerseits von der Qualität der Dienste ab und andererseits von der Qualität und Quantität der von den Projekten bereitgestellten Informationen. Um den Weg vom Problem zur Person, die das Problem lösen kann, zu verkürzen, wurde in Math&Industry der Dienst „Expertendatenbank“ entwickelt.

## 3 Die Entwicklung der Expertendatenbank

Zum genaueren Verständnis der Lösung und ihrer Entstehung, sei hier der Weg von der Konzeption zur Realisierung der Expertendatenbank nachgezeichnet.

### 3.1 Beschreibung des Konzepts

#### A) Grundlegenden Anforderungen

##### i. Interne Experten

Die Mitarbeiter der Projekte des BMBF-Mathematikprogramms sollen über die Expertendatenbank nachgewiesen werden. Sie werden im Folgenden auch als „interne Experten“ bezeichnet.

##### ii. Externe Experten

Es sollen zusätzlich weitere Experten in die Datenbank aufgenommen werden können (= „externe Experten“).

##### iii. Zugang: Web-Schnittstelle

Die Expertensuche soll über eine Web-Schnittstelle erfolgen.

##### iv. Generelle Anforderung: Einfachheit

Die Eintragung in die Expertendatenbank soll mit möglichst wenig Aufwand verbunden sein.

Ebenso ist ein nutzerfreundlicher (intuitiver) Zugang zur Expertendatenbank Voraussetzung für breite Akzeptanz.

Dazu müssen insbesondere die Expertisegebiete in geeigneter Form systematisiert werden.

## **B) Vorentscheidungen / Präferierte „Lösungspfade“**

**Zu i. Interne Experten:** In den Websites der Projekte gibt es Homepages der Projektmitarbeiter. Die Angaben dieser Homepages sollen für die Expertendatenbank ausgewertet werden. Da nicht davon ausgegangen werden kann, dass jeder in die Datenbank aufgenommen werden möchte, muss die Zustimmung dazu individuell erfragt werden.

**Zu ii. Externe Experten:** Externe Experten sollen sich in ähnlicher Weise wie interne Experten eintragen können.

**Zu iii. Zugang:** Web-Schnittstelle: Es soll eine Stichwortsuche sowie ein Browsen in den Expertenkategorien möglich sein.

**Zu iv. Generelle Anforderung:** Einfachheit: Für interne Experten sollten die Angaben der Mitarbeiterhomepages der Projektwebsites genutzt werden. Für die externen Experten muss jedoch eine separate Dateneingabe (als Web-Schnittstelle) möglich sein.

Weiterhin wurde gewünscht, dass auch für die Projektmitarbeiter ein individueller Web-Zugang möglich ist (statt des Projektzugangs über den WebSiteMaker).

Das Schema der Expertisegebiete soll nur wenige, grobe Kategorien (auf der Top-ebene) enthalten (nicht mehr als 15) und sich einerseits an den Themen der bisherigen Förderprogramme orientieren, andererseits aber allgemein genug gehalten sein, um nicht nach kurzer Zeit das Schema wieder ändern zu müssen. Die Experten sollen ihre Expertisegebiete durch Stichworte dezidiert benennen können. Diese Stichworte bilden dann sozusagen die Untergebiete.

## **C) Lösungsansätze**

**Zu iii. Zugang: Web-Schnittstelle:** Um ein effizientes Suchen (vor allem durch „Browsen“) zu ermöglichen, sollten alle Angaben die für Suche und Darstellung des Suchergebnisses nötig sind über eine Datenbank zur Verfügung stehen und nicht durch Parsen des RDF-Codes der Homepages erfolgen.

Die Expertendatenbank bietet eine Suche über die:

- Expertisegebiete (Kategorien plus Stichworte der Experten)

Als Ergebnis der Suche werden die Experten eines Gebietes angezeigt, und zwar insbesondere

- der Nachname und der Vorname
- sowie ein Link zu weiterführenden Informationen über den Experten (Homepage)

**Zu i. Interne Experten:** Wegen der Datenbankanforderung von C.iii müssen Daten der Projekthomepages auch in einer Datenbank vorgehalten werden. Dieses

kann durch direktes Schreiben der Daten in die Datenbank geschehen, oder durch ein separates Programm, dass in regelmäßigen Abständen (oder wenn das Modul zur Homepageerstellung (MIPMPers) beendet wurde), die Informationen in der RDF-Datei mit den Informationen in der Datenbank abgleicht.

**Zu ii. Externe Experten:** Durch die Ähnlichkeit der Aufgabe bietet es sich an, auch für die externen Experten bei der Dateneingabe die MIPMPers-Komponente des WebSiteMakers zu verwenden.

**Zu iv. Generelle Anforderung:** Einfachheit: MIPMPers ermöglicht Expertiseausagen und kann auch für die externen Experten eingesetzt werden. Eine separate, „projektlose“ Homepage kann für die internen Experten analog erstellt werden.

Das Klassifikationsschema der Expertisegebiete findet sich unter

<http://www.iwi-iuk.org/material/RDF/Schema/Class/expareas.html>

(Ein weit differenzierteres Vorgängerschema war

<http://www.iwi-iuk.org/material/RDF/Schema/Class/expf.html>

Es wurde jedoch als zu unflexibel und zu unübersichtlich verworfen.)

#### **D) Offene Probleme**

**Zu i. Interne Experten:** Eine Person kann mehrere Homepages haben (durch Mitarbeit in mehreren Projekten, möglicherweise ergänzt durch die Homepage eines individuellen Web-Zugangs). Dadurch können sich widersprechende Angaben auf den Homepages finden. Welche Angaben sollen dann in die Expertendatenbank einfließen, d.h. als Suchgrundlage dienen?

**Zu iii. Zugang: Web-Schnittstelle:** Eine Person kann mehrere Homepages haben. Auf welche soll das Suchergebnis verweisen, nachdem eine Suchanfrage einen Treffer ermittelt hat? Das Auswahlproblem könnte umgangen werden, indem auf alle vorhandenen Homepages verwiesen wird.

#### **E) Auswahlprinzip als Lösung der offenen Probleme**

**Zu i. Interne Experten:** Jede Homepage erzeugt einen kompletten Datensatz in der Datenbank (statt eine Datenmodellierung zu wählen, die die Relation „1 Person besitzt n Homepages“ redundanzfrei abbildet). So wird verhindert, dass das Ändern der Daten einer Homepage, Daten einer anderen ebenfalls verändert.

**Zu iii. Zugang: Web-Schnittstelle:** Durch E.i wurde das Konsistenzproblem in ein Auswahlproblem überführt, welches lautet: Welche Homepage enthält die Daten, die für die Expertendatenbank relevant sind? Folgende Heuristik löst den Konflikt:

- 1) Existiert eine Homepage mit einem individuellen Login (statt eines Projektzugangs) so werden die Daten dieser Homepage genommen. Denn was eine Per-

son mit Sicherheit eigenhändig über sich aussagt, hat Vorrang vor dem, was möglicherweise ein Webbeauftragter eines Projekts über jemanden anders ausgesagt hat.

- 2) Existieren mehrere projektbezogene Homepages, so werden die Daten des jüngsten Projekts genommen.

### 3.2 Das Ergebnis

Nach Lösung aller konzeptionellen Probleme, ist mit der Expertendatenbank ein Dienst entstanden, der auf eine einfache Art und Weise von einer Fragestellung zu Menschen führt, die fähig und bereit sind Antworten zu geben bzw. zu finden.

**Expertendatenbank**

Willkommen bei der Expertendatenbank der Angewandten Mathematik  
- ein Math&Industry Service

Die Expertendatenbank gibt Auskunft über Wissenschaftler, die in den letzten Jahren konkrete Probleme aus der Industrie und dem Dienstleistungsbereich mit mathematischen Mitteln und Methoden bearbeitet und gelöst haben. Die Expertendatenbank listet diese Personen und ihre Expertisegebiete auf.

Sie suchen einen Experten aus dem Bereich:	Unsere Experten und Ihre Expertisegebiete:
<ul style="list-style-type: none"> <li>◆ Energietechnik, -erzeugung, -verteilung (1)</li> <li>◆ Entwurf und Konstruktion von Mechanischen Systemen (1)</li> <li>◆ Mikro- und Optoelektronik, Chipdesign (5)</li> <li>◆ Nanotechnologie, Mikrosystemtechnik (2)</li> <li>◆ Telekommunikation, Internettechnologien (5)</li> <li>◆ Verkehr, Transport, Logistik (3)</li> <li>◆ ... Sonstige Anwendungsbereiche (1)</li> </ul>	<p><b>Ausgewählt: Verkehr, Transport, Logistik</b></p> <p>Bomdörfer, Ralf    Verkehr, Transport, Logistik (Mathematische Optimierung)</p> <p>Grötschel, Martin    Telekommunikation, Internettechnologien (Öffentlicher Personennahverkehr) Entwurf und Konstruktion von Mechanischen Systemen (Entwurf von Telekommunikationsnetzen) Verkehr, Transport, Logistik (Entwurf von Telekommunikationsnetzen, Funknetzplanung (UMTS, GSM)) Nanotechnologie, Mikrosystemtechnik (Halbleiterentwurf)</p> <p>Römisch, Werner    Mikro- und Optoelektronik, Chipdesign (Rauschanalyse in der Schaltungssimulation) Verkehr, Transport, Logistik (Optimierung der Ertragssteuerung von Flugbuchungsanfragen) Energietechnik, -erzeugung, -verteilung (Optimierung in der Energiewirtschaft)</p>

Alle Experten

Stichwort(e) (Expertisegebiete)

☒ Groß- und Kleinschreibung ignorieren

(Für die Mitglieder der Expertendatenbank gibt es eine Webschnittstelle.)

W3C HTML 4.01

Abbildung 2: Suchschnittstelle der Expertendatenbank

([http://www.mathematik-21.de/cgi-bin/queryexdb\\_dyn.cgi](http://www.mathematik-21.de/cgi-bin/queryexdb_dyn.cgi))

#### 3.2.1 Suche

Auf der Seite der Suchschnittstelle der Expertendatenbank ([http://www.mathematik-21.de/cgi-bin/mathInd/queryexdb\\_dyn.cgi](http://www.mathematik-21.de/cgi-bin/mathInd/queryexdb_dyn.cgi)) lässt sich gut sehen, wie das „Browsen“ realisiert ist (siehe Abbildung 2). Auf der linken Seite finden sich die Anwendungsbereiche, denen sich Experten zugeordnet haben. Wird dort ein Bereich ausgewählt, so sind in der linken Spalte die entsprechenden Experten mit allen ihren Expertisegebieten zu sehen, einschließlich den Untergebieten (= den persönlichen Stichworten) dazu. Der Name eines Experten ist mit seiner Homepage verlinkt. Zusätzlich ist es möglich sich alle Experten anzeigen zu lassen oder nach Stichwor-

ten zu suchen. In letzterem Fall werden die Namen der Gebiete und Untergebiete durchsucht.

### 3.2.2 Eingabe

Die Gestaltung des Web-Zugangs zu den individuellen Seiten ist schlicht gehalten (<http://www.mathematik-21.de/cgi-bin/mathInd/loginexdb.cgi>): es finden sich dort zwei Felder für Login und Passwort; nach erfolgter Anmeldung wird MIPMPers gestartet mit dem sich dann die eigenen Daten ändern lassen.

Auf den WebSiteMaker für den Projektzugang wird hier nicht weiter eingegangen. Zum WebSiteMaker siehe (Grötschel et al. 2005).

Das Formular zur Erzeugung einer Homepage (MIPMPers) enthält viele vordefinierte Felder und bietet somit zahlreiche Möglichkeiten Informationen strukturiert darzustellen. Letztendlich ist die Anzahl der Informationen unbeschränkt die hier untergebracht werden können (von technischen Limitierungen abgesehen), da es ein Feld gibt, in dem beschreibender Text zur Homepage eingegeben werden kann – einschließlich der Möglichkeit (X)HTML zur Gestaltung zu verwenden.

Forschung / Kenntnisse / Erfahrungen in Anwendungsgebieten

1. Bitte wählen Sie ein Expertisegebiet aus:

Oder geben Sie ein anderes Gebiet an: (z.B. Reaktorsteuerung, Verkehrsplanung, ...)

2. Ordnen Sie Ihr Expertisegebiet einem der folgenden Bereiche zu:

Hier können Sie einen zusätzlichen Bereich vorschlagen:

☐ Ich stehe für Anfragen bzgl. oben genannter Anwendungsgebiete zur Verfügung (Sie werden dann in die Math&Industry Expertendatenbank aufgenommen.)

Abbildung 3: Formularausschnitt zur Angabe der Kenntnisse in Anwendungsgebieten

Der Bereich des Formulars, der die Kenntnisse in Anwendungsgebieten erfragt, ist als eine 2 x 2 Matrix gestaltet, die genau zwei Eingaben verlangt (siehe Abbildung 3). In der Zeile 1 soll frei angegeben werden, worin sich diejenige Person besonders gut auskennt. Um eine erhöhte Standardisierung der verwendeten Begriffe zu ermöglichen, kann in dem ersten Feld aus den bereits in der Datenbank vorhandenen Begriffen einer ausgewählt werden. Wenn kein passender dabei ist, so steht rechts daneben ein Freitextfeld für ein eigenes / neues Arbeitsgebiet zur Verfügung. Die 2. Zeile dient der Zuordnung des in der 1. Zeile genannten Gebiets zu einem der 13 Bereiche aus dem erarbeiteten Expertiseschema. Falls keiner dieser Bereiche passt, so besteht wiederum die Möglichkeit eine eigene Angabe in einem Freitextfeld zu machen. Das in der 1. Zeile genannte Gebiet wird dann dem Bereich „Sonstige Be-

reiche“ zugeordnet und Math&Industry wird automatisch per Mail von diesem neuen Bereichsvorschlag informiert. Nach Begutachtung, kann dieser Bereich später noch dem Schema zugefügt und das Gebiet aus der Kategorie „Sonstige Bereiche“ der neuen Kategorie zugewiesen werden.

### *3.2.3 Ausgabe*

Als letzter Schritt wird dann noch die Homepage erzeugt, einschließlich ihrer Beschreibung in RDF/XML.

## **4 Fazit zu Math&Industry**

Mit dem Portal von Math&Industry und vor allem der Expertendatenbank ist die Grundlage für ein bedeutendes Stück Vernetzung der mathematischen Wissenschaften und Öffentlichkeits- / Lobbyarbeit für die Mathematik als Ganzes gelegt worden.

Probleme die noch zu klären sind, sind die Motivierung der Projekte, ihre Arbeit im Sinne von Math&Industry aufzubereiten (und zwar schon mit Beginn des Projekts) und die Absicherung der Expertendatenbank vor Missbrauch. Die Wahrscheinlichkeit, dass sich Personen als Experten eintragen, die keine sind, ist zwar gering, da nur Projektmitarbeiter auch einen Zugang zum WebSiteMaker (und damit die Möglichkeit zum Erstellen beliebiger Homepages) haben. Doch lassen sich vorsätzliche Falschaussagen nicht ausschließen. Daher ist geplant in nächster Zeit eine Arbeitsgruppe zu bilden, die bei den Expertenzuordnungen ein Vetorecht haben soll und die Motivationsaufgabe übernimmt.

Weiterhin steht eine Erweiterung auf weitere Förderinstitutionen einschließlich einer Internationalisierung noch aus. Ebenso wie die Abstrahierung des Konzepts. Dadurch können die Struktur und die Software-Werkzeuge von Math&Industry auch auf andere Wissenschaften angewendet werden. Ein erstes Beispiel dafür ist das Programm „Netzwerke Grundlagenforschung erneuerbare Energien und rationale Energieanwendungen“<sup>11</sup>.

## **5 Literaturangaben**

Böckmann, C.; Wauer, J. (2001): Algorithms for the inversion of light scattering data from uniform and non-uniform particles. *Journal of Aerosol Science*. 32, S. 49-61.

---

<sup>11</sup> <http://www.fz-juelich.de/ptj/netzwerke-grundlagenforschung/>, die Freigabe des entsprechenden Portals wird voraussichtlich im März 2007 erfolgt sein.

- Eisenblätter, Andreas; Fügenschuh, Armin; Koch, Thorsten; Koster, Arie; Martin, Alexander; Pfender, Tobias; Wegel, Oliver; Wessäly, Roland (2002): Modelling Feasible Network Configurations for UMTS. In: Telecommunications Network Design and Management. Anandalingam, G.; Raghavan, S. (Hrsg.) Boston: Kluwer Academic Publ. S. 1-24.
- Gardner, R.; Gritzmam, P. (1999): Uniqueness and complexity issues in discrete tomography. In: Herman, G.; Kuba, A. (Hrsg.): Discrete Tomography: Foundations, Algorithms, Applications. Boston: Birkhauser.
- Grötschel, Martin; Roggenbuck, Robert; Sperber, Wolfram (2005): Entwicklung und Aufbau des Internet-Dienstes „Math&Industry“ – Abschlußbericht zur 3. Förderperiode (2001-2004). In: ZIB-Report 05-23.
- Herrmann, Norbert (2005): Mathematik ist überall. Mathematik im Alltag / alltägliche Mathematik. München, Wien: Oldenbourg.
- Löbel, Andreas (1999): Solving Large-Scale Multiple-Depot Vehicle Scheduling Problems. In: Proceedings of the 7<sup>th</sup> International Workshop on Computer-Aided Transit Scheduling, Berlin, S. 193-220.
- Roggenbuck, Robert (2005): Glossare als Informationsdienste – Das Beispiel Math&Industry. Vortrag auf der 11. IuK Jahrestagung 2005, 10.5.2005, Bonn – Bad Godesberg, Vortragsfolien: [http://www.mathematik-21.de/publications/050509IuK\\_Bonn/050509IuK-Glossare.pdf](http://www.mathematik-21.de/publications/050509IuK_Bonn/050509IuK-Glossare.pdf)
- Roggenbuck, Robert; Sperber, Wolfram (2006): CMS fürs Semantic Web: eine Analyse anhand der Erfahrungen des Math&Industry-Projekts. In: Eckstein, Rainer; Tolksdorf, Robert (Hrsg.): XML Tage 2006 in Berlin – Tagungsband, Humboldt-Universität zu Berlin.
- Roggenbuck, Robert; Sperber, Wolfram (2005): Meta data Driven Website Generation – Tools for Describing Projects. Vortrag auf dem Workshop „Preservation and DC-Tools: Standards and Standardisation activities“, 1.3.2005, Vortragsfolien: [http://www.mathematik-21.de/publications/050301cashmereGoettingen/roggenbuck\\_metadata\\_generation.pdf](http://www.mathematik-21.de/publications/050301cashmereGoettingen/roggenbuck_metadata_generation.pdf)



# Change Management und Strukturveränderung bei IuK-Projekten an Universitäten\*

*Susanne Pillay*

IT-Servicezentrum , Universität Augsburg  
Universitätsstraße 8, D-86159 Augsburg  
*susanne.pillay@its.uni-augsburg.de*

## **Zusammenfassung**

Integriertes Informationsmanagement gewinnt zunehmende Bedeutung für Hochschulen, um den gegenwärtigen Herausforderungen zu begegnen. Die Universität Augsburg hat sich daher die Aufgabe gesetzt, ein IT-Servicezentrum (ITS) einzurichten, unter dessen Dach die IuK-Dienstleistungen für Forschung, Lehre, Studium und Verwaltung in einem Kooperationsmodell gebündelt werden. Im folgenden Beitrag werden zunächst Ausgangslage und Arbeitsfelder des ITS-Projektes vorgestellt. Im Weiteren stehen dann die Gestaltung dieses umfassenden und dynamischen Veränderungsprozesses mit dem Ziel einer nachhaltigen Strukturveränderung und die konkreten Erfahrungen an der Universität Augsburg im Vordergrund.

## **Abstract**

Integrated Information Management has become important in fulfilling requirements which face Higher Institutions of Education today. Due to this challenge, the University of Augsburg has taken steps in implementing an IT-Service Center (ITS) on its campus. The Information and Communication Services for Research, Teaching, Studies and Administration shall be supplied through the cooperation model ITS. The following article begins with an introduction of the initial positions and the field of work, of the ITS Projects. It proceeds further to explain the configuration of the extensive and dynamic change processes with the objective of attaining sustainable structural changes and to crystallize the concrete experiences made by the University of Augsburg.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 445-463.

## I Projektdarstellung

### 1.1 Universitäre IuK-Versorgung vor neuen Herausforderungen

Der internationale Wettbewerb der Hochschulen, die Studienreform (Bologna, Studiengebühren), wachsende Studierendenzahlen, interuniversitäre Kooperationen sowie der gewohnte Umgang mit digitalen Medien bei der gegenwärtigen Studierendengeneration beschleunigen den Einzug von Informations- und Kommunikationstechnologien (IuK) in alle Bereiche der Universität. Dies stellt die Universitäten derzeit vor neue Anforderungen an ihre IuK-Versorgung für Forschung und Lehre bzw. Studium ebenso wie für die Verwaltung. Dem stehen auf Seiten der IuK-Anbieter einer Universität über Jahrzehnte gewachsene uneinheitliche Strukturen, unterschiedliche Systemlandschaften und Ineffizienzen in der Erbringung der IuK-Leistungen gegenüber. Abbildung 1 zeigt die Anforderungen, die sich aus diesen Rahmenbedingungen auf Nutzer- und Anbieterseite für eine integrierte IuK-Versorgung ergeben und wie dieser Herausforderung an der Universität Augsburg durch die Einrichtung eines IT-Servicezentrums begegnet wird.

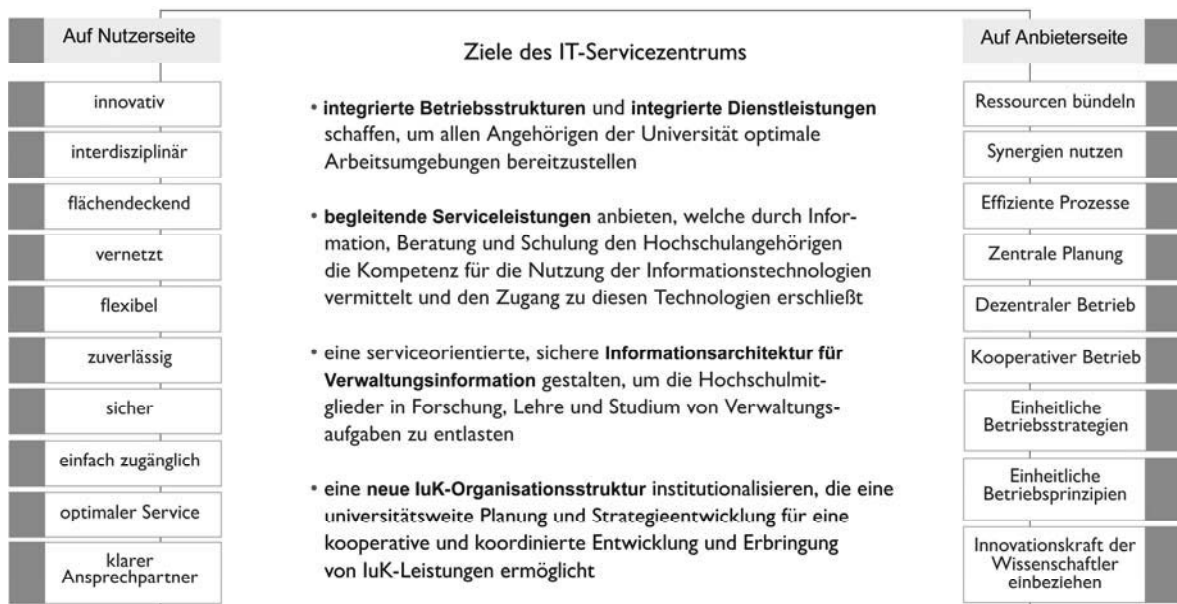


Abbildung 1: Anforderungen an die integrierte IuK-Versorgung an der Universität Augsburg

### 1.2 Universität Augsburg und ihre IuK-Dienstleistungen im Profil

Die Studierendenzahlen an den sieben Fakultäten der Universität Augsburg liegen derzeit bei 14.800 Studentinnen und Studenten, die sich zu etwa 50% auf die Geistes- und Sozialwissenschaften mit dem Schwerpunkt der Lehrerbildung, zu 35% auf

Wirtschaftswissenschaften und Jura und zu 15 % auf die Naturwissenschaften und Informatik verteilen. Neben ca. 160 Professoren und 550 wissenschaftlichen Mitarbeitern sind rund 550 Mitarbeiter im wissenschaftsstützenden Bereich an der Universität Augsburg beschäftigt. Die IuK-Dienstleistungen werden bisher als Summe von Einzelleistungen verschiedener Einrichtungen und Einzelpersonen erbracht. Dazu zählen:

- die fünf voneinander unabhängigen Infrastruktureinrichtungen:
- Universitätsbibliothek, Rechenzentrum, Videolabor sowie Abteilung Bau und Technik und DV-Gruppe der zentralen Universitätsverwaltung
- die DV-Betreuer der Fakultäten und Zentralen Einrichtungen
- das Referat für Presse- und Öffentlichkeitsarbeit des Rektorats
- einzelne wissenschaftliche und wissenschaftsstützende Mitarbeiter direkt an den Lehrstühlen und in den Einrichtungen

Außer der Zusammenarbeit im Rahmen der geregelten Aufgabenteilung zwischen dem Rechenzentrum – für den Betrieb der hochschulweiten IT-Infrastruktur – und den DV-Betreuern der Fakultäten und Zentralen Einrichtungen – für die dezentrale DV-Grundversorgung –, hat es eine Abstimmung quer über alle Einrichtungen hinweg bisher nur in besonderen Einzelfällen gegeben (vgl. Eichner u a. 2006). Mit dem Aufbau eines IT-Servicezentrums verfolgt die Universität nun das Ziel, alle an der Erstellung von IuK-Dienstleistungen beteiligten Einrichtungen und Personen unter der Leitung eines CIO-Gremiums zu einer durch eine IuK-Gesamtstrategie aufeinander abgestimmten und gemeinsam gestalteten Leistungserstellung zusammenzuführen. Diese soll neben effizientem Ressourceneinsatz auf Seiten der IuK-Anbieter insbesondere die konkreten Anforderungen der Nutzer von IuK-Leistungen im Fokus haben und Leaduser aktiv in die Entwicklung neuer Dienste und Serviceleistungen mit einbeziehen (vgl. Sporer u a. 2007).

### **1.3 Aufbau des IT-Servicezentrums**

Die Universität Augsburg wird bei der Neugestaltung ihrer IuK-Organisationsstruktur mit ihrem Projektantrag zum „Aufbau eines IT-Servicezentrums“ im Rahmen der Innovationsinitiative „Leistungszentren für Forschungsinformation – Integriertes Informationsmanagement“ seit Januar 2006 von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Vor der Erstellung des Projektantrags wurde die Ausgangssituation im IuK-Bereich der Universität Augsburg analysiert, woraus

sich für ein umfassendes Konzept zum Integrierten Informationsmanagement im Wesentlichen zwei große Aufgabenbereiche mit vier Arbeitsfeldern (AF) ergaben:

- 1) die nachhaltige Implementierung einer neuen IuK-Organisationsstruktur (AF Struktur), in Form einer institutionalisierten und engen Zusammenarbeit der an der Erstellung der IuK-Dienste Beteiligten, sowie
- 2) der Aufbau und die Integration von technischer Infrastruktur (AF Technik) mit den dazugehörigen Veränderungen in den Abläufen (AF Prozesse) und bei den Universitätsangehörigen (AF Person).

### *1.3.1 Implementierung der neuen IuK-Organisationsstruktur*

Das Augsburger Modell für eine neue IuK-Organisationsstruktur sieht mit dem IT-Servicezentrum die Schaffung einer übergeordneten Institution vor, die alle IuK-Dienstleister und -Nutzer als Kooperationspartner unter ihrem Dach zusammenführt und koordiniert. Dabei wird von einer organisatorischen Zusammenlegung der IuK-Anbieter in einer Einrichtung (Zentralisierung) abgesehen. Vielmehr sollen die gegenwärtigen Kernkompetenzen der einzelnen Dienstleister erhalten bleiben und durch gezielte Abstimmung Ressourcen gebündelt und Synergien genutzt werden (Kooperation). Als zugrundeliegende Prinzipien der neuen Organisation gelten:

1. Hochschulweite Zuständigkeit des IT-Servicezentrums für alle Themen aus dem IuK-Bereich und Etablierung des ITS als zentraler IuK-Ansprech- und Kooperationspartner durch Geschäftsstelle und Kundenbüro
2. Zentrale Leitung des ITS und zentrale Koordination aller Beteiligten durch Lenkungsrat (CIO-Gremium) und Geschäftsstelle mit dem Fokus auf gemeinsame Planung, Strategie und Ressourcenbündelung
3. Bereichsübergreifende Zusammenarbeit aller IuK-Mitarbeiter unter Einbeziehung von Leadusern, um IuK-Angebote kooperativ in Projekt- und Arbeitsgruppen neu zu entwickeln bzw. in Servicegruppen mit einheitlichen Betriebs- und Servicestrategien zur Verfügung zu stellen

Abbildung 2 zeigt die sich daraus ergebenden Strukturelemente des IT-Servicezentrums. Ziel ist es, gemeinsam eine nachhaltige Struktur und Abläufe zu gestalten, die langfristig integriertes Informationsmanagement in allen Bereichen der Universität ermöglichen. Dabei sollen im Sinne eines „sowohl als auch“ die Vorteile der dezentralen Leistungserbringung – wie Nähe zum Bedarf der jeweiligen Kunden und Flexibilität – mit den Vorzügen einer zentralen Klammer – wie die Aus-

richtung auf eine gemeinsame Strategie und auf geteilte Grundsätze und Prinzipien – verbunden werden (vgl. auch Doppler u a. 2002, S. 23).

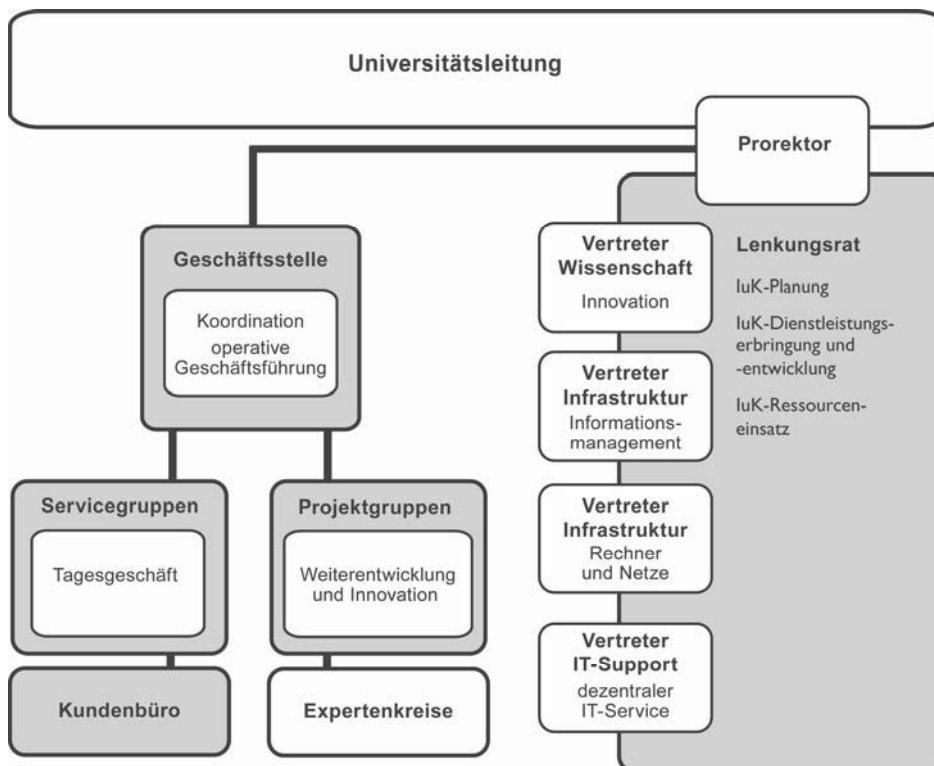


Abbildung 2: Das IT-Servicezentrum an der Universität Augsburg

Bezogen auf das Thema Change Management ist insbesondere für das Verständnis des folgenden Beitrags erwähnenswert<sup>1</sup>: Die ITS-Geschäftsstelle ist neben der Koordination aller Beteiligten für das Projekt- und Change-Management zuständig. Im Zuge der Reorganisation der Arbeitsverteilung im IuK-Bereich, werden zukünftig die Dienste und Support-Leistungen nicht mehr von den einzelnen Einrichtungen jeweils für ihren eigenen Bereich erbracht, sondern von kooperativen Servicegruppen, die vom Lenkungsrat aufgabenbezogen aus Experten der verschiedenen IuK-Einrichtungen gebildet werden. Für den Aufbau neuer und die Integration bestehender technischer Infrastrukturen und die damit zusammenhängenden prozess- und personen-bezogenen Veränderungen sowie für die Weiterentwicklung bestehender und die Entwicklung innovativer neuer IuK-Dienst- und Serviceleistungen werden vom Lenkungsrat temporär und bedarfsorientiert Projektgruppen eingesetzt. In die Projektarbeit sollen insbesondere auch Wissenschaftler und Studierende einbezogen werden (vgl. Sporer u a. 2007). Bei Bedarf können für kurzfristige

<sup>1</sup> Ausführlichere Informationen zu den Strukturelementen sowie zu den Inhalten und Zielen der unter 1.3.2 genannten Projekte sind zu finden unter: [www.uni-augsburg.de/its](http://www.uni-augsburg.de/its)

und innerhalb eines klar definierten Rahmens zu erbringende Arbeitsaufträge von den Projektgruppen zusätzlich Arbeitsgruppen gebildet werden.

### 1.3.2 Aktuelle Projekte des IT-Servicezentrums

Für die nachhaltige Implementierung der neuen IuK-Organisationsstruktur und entsprechender kooperativer Prozessabläufe wurde vom Lenkungsrat das Strukturprojekt eingerichtet. Zudem wurden zu den drei Arbeitsschwerpunkten *Leistungsfähige IT-Basis*, *Serviceorientierte Verwaltung* sowie *Forschung und Lehre/Lernen* neun Teilprojekte etabliert. Sie verfolgen sowohl technische als auch organisatorische Ziele und erarbeiten für die neu geschaffene Infrastruktur ergänzende Serviceleistungen wie Informations-, Beratungs- und Schulungsangebote zur bedarfsgerechten Begleitung der Nutzer bei der Anwendung der technischen Lösungen. Dadurch wird die Grundlage gelegt, dass Wissenschaftler und Studierende sich auf ihre eigentlichen Aufgaben konzentrieren können, indem

- im Schwerpunkt *Leistungsfähige IT-Basis* ein umfassendes, bedarfsgerechtes und zuverlässiges Angebot an IuK-Diensten und Serviceleistungen für die Benutzerarbeitsplätze zur Verfügung gestellt wird,
- durch die Teilprojekte des Arbeitsbereiches *Serviceorientierte Verwaltung* die Verwaltungstätigkeiten auf elektronischem Weg zukünftig schneller und einfacher erledigt werden können,
- der Schwerpunkt *Forschung und Lehre/Lernen* den kompetenten Medieneinsatz fördert und integrierte Informationsangebote bereitstellt.

Gemeinsam ist allen Teilprojekten ihre systematische Vorgehensweise: Ausgehend von einer umfassenden Analyse der Ausgangssituation (d.h. bestehender Systeme, Prozesse, Strukturen, Betriebsstrategien, Kooperationen etc.) erfolgt eine umfassende Identifikation des Bedarfs von Nutzern und IuK-Mitarbeitern sowie der technischen Anforderungen an neue Lösungen, um schließlich innovative, ressourcenschonende und bedarfsgerechte Konzepte und Modelle für neue IuK-Angebote entwickeln zu können. Zudem arbeiten die Teilprojekte themenorientiert innerhalb und zwischen den drei Bereichen eng zusammen und stimmen sich über ihre Aktivitäten ab, um integrierte Lösungen zu schaffen und dabei alle Betroffenen mit einzubeziehen.

Wie diese kurze Beschreibung des ITS-Projektantrags zeigt, setzt die erfolgreiche Bewältigung umfassende Veränderungen in der Universität voraus, d.h. eine Neugestaltung von Strukturen, Systemen, Prozessen und Arbeitsweisen. Der Übergang von der ursprünglichen Organisation zur neuen ist ein komplexer Change-Prozess,

der Begleitung in Form eines professionellen Change-Managements benötigt. Dies soll im Folgenden am Beispiel des ITS näher ausgeführt werden.

## **2 Change-Management im ITS-Projekt**

### **2.1 Veränderungen im IuK-Bereich als umfassendes Change-Projekt**

Einzelne Veränderungen im IuK-Bereich, wie die Einführung bzw. Ablösung eines Systems oder das Angebot einer neuen Dienstleistung, gehörten bereits vor Projektbeginn zum Alltag der IuK-Anbieter. Das Besondere am ITS-Projekt ist nun, dass es bereichsübergreifend angelegt ist und hochschulweite Auswirkungen hat, sowohl bei den IuK-Anbietern wie auch bei den Nutzern. Es handelt sich um einen geplanten und koordinierten Veränderungsprozess, der den gesamten IuK-Bereich der Universität Augsburg einbezieht. Es geht nicht nur darum, durch Optimierungsmaßnahmen innerhalb einer Einrichtung oder eines Tätigkeitsfeldes *besser* zu werden, sondern durch bereichsübergreifende Kooperationen und die Konzeptionierung innovativer Angebote im Ganzen *anders* zu werden im Sinne einer „innovativen Transformation“ (vgl. Heitger u a. 2002, S. 41). Ein Veränderungsprojekt dieser Art steht vor einer doppelten Herausforderung: Einerseits handelt es sich um ein inhaltliches Forschungs- und Entwicklungs-Projekt (F&E) und andererseits um ein Changevorhaben auf drei Ebenen, d.h. mit technischen, organisatorischen und personenbezogenen Veränderungen. Sowohl die Beratungsliteratur wie auch die Erfahrungen im ITS-Projekt zeigen, dass für den Umgang mit einem solch umfassenden Veränderungsprozess ein professionelles Change-Management notwendig ist. Aufgabe des Change-Managers ist es dabei, (1) die Beteiligten im Veränderungsprozess zu begleiten, um den gegenwärtigen Veränderungsbedarf zu bewältigen, und (2) gleichzeitig den Aufbau einer eigenen Veränderungskompetenz innerhalb der Organisation zu fördern.

Die Begleitung der Organisation, d.h. Prozessunterstützung zu leisten, bedeutet für den Change-Manager u a.: (1) den Veränderungsprozess als solchen zu thematisieren und die Aufmerksamkeit gleichermaßen auf alle drei Ebenen des Prozesses zu lenken, (2) alle Betroffenen in den Veränderungsprozess einzubinden und ihnen Unterstützung im individuellen Umgang mit den Veränderungen anzubieten, (3) geeignete Strukturen und aufeinander abgestimmte Maßnahmen für diesen komplexen Veränderungsprozess zu entwerfen und ihre Umsetzung zu steuern sowie (4) einen Raum in der Organisation zu schaffen, in dem eine angstfreie Veränderung in Richtung von Kooperation, Integration und Innovation ermöglicht wird.

Zusätzlich zu dieser Prozessunterstützung hat der Change-Manager nach dem Prinzip der „Hilfe zur Selbsthilfe“ den Aufbau einer Change-Kompetenz innerhalb der Organisation zum Ziel: Dazu gehören die Bereitschaft und Kompetenz, die Veränderungsnotwendigkeiten zu erkennen und die Veränderungsfähigkeit des Einzelnen sowie der Organisation als Ganzes weiterzuentwickeln, und sich mit komplexen Veränderungsprozessen auseinanderzusetzen und sie professionell planen und steuern zu können (vgl. Heitger u a. 2002). Damit soll erreicht werden, dass die durchgeführten Veränderungen nachhaltig wirken und dass die Organisation und ihre Individuen befähigt werden, weiteren Veränderungsbedarf selbständig zu erkennen und entsprechende Changeprozesse proaktiv, von sich aus anzustoßen.

Der Fokus des professionellen Change-Managers muss gleichermaßen auf dem Changeprozess (Gestaltung und Steuerung) wie auf den Prozessbeteiligten (Kommunikation, Abstimmung, Qualifizierung und Partizipation am Prozess) liegen – immer in dem Bewusstsein, dass seine Möglichkeiten der Gestaltung und Steuerung eines komplexen und dynamischen Veränderungsprozesses begrenzt sind und dass die Beachtung der Wechselwirkung zwischen den Personen und dem Prozess ein wichtiger Erfolgsfaktor ist. Im Sinne einer Case-Study soll im Folgenden ausgehend von entsprechender Beratungsliteratur eine erste Aufarbeitung der Erfahrungen mit der Einbeziehung der Beteiligten und mit der Gestaltung und Steuerung des Change-Prozesses im ITS-Projekt vorgestellt werden.

## **2.2 Partizipation am Veränderungsprozess**

Zunächst soll der Fokus auf die Prozessbeteiligten gelegt werden. Für die Umsetzbarkeit der Konzepte und die Nachhaltigkeit der durchgeführten Veränderungen an der Universität Augsburg spielt die Akzeptanz auf Anbieter- und Nutzerseite sowohl für die spätere Anwendung der technischen Lösungen und die Weiterführung der entwickelten Strategien als auch im Umgang mit den optimierten Prozessen und kooperativen Strukturen eine große Rolle. Da Veränderungen in Organisationen letztlich immer persönliche Veränderungen der einzelnen Menschen bedingen, sollten sie durch entsprechende Maßnahmen zur Information, Motivation, Qualifizierung und Mitbestimmung der Betroffenen begleitet werden.

Im ITS-Projekt wird Akzeptanz insbesondere durch eine frühzeitige Partizipation der IuK-Mitarbeiter und -Nutzer geschaffen. Die „Betroffenen“ werden u a. durch folgende Maßnahmen zu „Beteiligten“ am Veränderungsprojekt gemacht: Sie werden über das Changevorhaben informiert und zu ihrem Bedarf befragt, sie können Feedback einbringen und haben die Möglichkeit, in den themenbezogenen Arbeits- und Projekt-



gruppen oder als Pilotanwender ihren zukünftigen IuK-Alltag mitzugestalten. Der Erfolg der gemeinsamen Projektarbeit wird dadurch unterstützt, dass Weiterbildungen zu Change- und Projekt-Management, Kreativitätsmethoden oder Moderationstechniken angeboten werden und dass in Workshops – neben der fachlichen Arbeit – der Fokus auch auf kulturelle Unterschiede in den Abteilungen gelegt wird und verschiedene Sichtweisen mit Hilfe eines Moderators strukturiert bearbeitet werden. Durch diese Maßnahmen kann einerseits vorgebeugt werden, dass bei den Betroffenen Widerstand entsteht, weil sie sich übergangen fühlen oder weil die geplanten Veränderungen in ihren Augen nicht sinnvoll bzw. nötig sind. Andererseits kann dadurch verhindert werden, dass Veränderungen am Bedarf der Betroffenen vorbeigehen, denn wie die Befragungen zeigen, sind zum Teil sehr heterogene Nutzeransprüche vorhanden, die ein differenziertes Angebot an IuK-Leistungen erfordern.

Als weiterer wichtiger Aspekt bei der Eröffnung von Mitbestimmungsmöglichkeiten hat sich die Klärung der Frage gezeigt, wer letztendlich Entscheidungen fällen kann. Hier wird offensichtlich, dass es genauso wichtig ist, wie die Partizipation überhaupt zu ermöglichen, die Grenzen der Mitbestimmung deutlich zu machen. Bildlich gesprochen müssen vom Change-Manager einerseits die Linien des Spielfeldes und die Spielregeln klar vorgegeben werden (Grenzen aufzeigen als schließende Maßnahme), andererseits sollte es dann genügend Raum für die Beteiligten geben, um das Spiel selbst frei auszugestalten (Freiraum als öffnende Maßnahme). Der Veränderungsprozess kann aus Sicht der Beteiligten viele Chancen, aber auch Gefahren bergen. Um ihnen dabei Sicherheit und eine klare Richtung zu geben, sollte Transparenz über die folgenden Punkte bestehen (siehe Tabelle 1):

Veränderungsrichtung und Inhalte des CV	• Was wollen wir verändern?
Notwendigkeit des CV	• Was soll so bleiben, wie es ist?
	• Warum müssen wir etwas verändern?
	• Was würde passieren, wenn wir nichts verändern?
Vorgehensweise des CV	• Wie wollen wir verändern (Change-Architektur etc.)?
Beteiligte am CV und ihre Mitgestaltungsmöglichkeiten und -grenzen	• Wer soll mit welchem Know-how (Fachwissen, Methodenkompetenz) am Veränderungsprozess in welcher Weise (Diskussion, Empfehlung, Expertenberatung, Entscheidung) beteiligt sein?

*Tabelle 1: Transparenz im IuK-Changevorhaben (CV)*

## 2.3 Gestaltung des Veränderungsprozesses

Für die professionelle Gestaltung des Veränderungsprozesses wurde im ITS-Projekt folgende – aus der systemischen Organisationsentwicklung entlehnte (vgl. z. B. Königswieser u a. 2004, S. 46) – Vorgehensweise gewählt: (1) Analyse, (2) Grobplanung, (3) Feinplanung, (4) Umsetzung der Maßnahmen sowie (5) Evaluation und Anpassung der Planung. Dabei wurde ausgehend von einer ausführlichen Diagnose der Ausgangssituation ein aufeinander abgestimmtes Maßnahmenbündel geplant. Nach der Durchführung einzelner Maßnahmen wurden ihre Wirkungen auf das Gesamtsystem evaluiert und die weiteren geplanten Maßnahmen entsprechend angepasst bzw. geeignete zusätzliche Maßnahmen entwickelt. Diese flexible Steuerung des dynamischen Changeprozesses trägt der permanenten „Veränderung der Veränderung“ (Heitger u a. 2002, S. 20) Rechnung.

### 2.3.1 Projekt- und Projektumfeldanalyse

Nach Königswieser (2004) ist vor Beginn des Veränderungsprozesses eine umfassende Diagnose der Situation durchzuführen. Dabei sollen u a. die angestrebten Ziele des Change-Projektes (d.h. technische, organisatorische und personenbezogene Veränderungen) und seine Beteiligten (d.h. ihre Interessenlage, Rollen, Arbeitskultur und Denkmuster) sowie das Projektumfeld (d.h. Stellenwert des Projektes in der Organisation; Informationsstand und Akzeptanz bei relevanten Gruppen) analysiert werden (vgl. Königswieser u a. 2004, S. 57). Dabei zeigten sich für das ITS-Projekt u a. folgende Ergebnisse:

Die inhaltliche Arbeit im Rahmen des F&E-Projektes steht mit technischen Verbesserungen (Modernisierung der Infrastruktur, Integration der Daten- und Systemlandschaft, Angebot innovativer Dienste etc.) sowie organisatorischen Veränderungen (Analyse und Optimierung der Prozesse und Strukturen zur IuK-Leistungserbringung) in der Perspektive der Beteiligten im Vordergrund; der damit einhergehende Changeprozess und die personale Ebene werden weitestgehend nicht bedacht, obwohl die erfolgreiche Umsetzung des technischen und organisatorischen Aspekts dieses Veränderungsprozesses letztlich davon abhängen, inwieweit es gelingt, die betroffenen Personen in den Planungs- und Implementierungsprozess mit einzubeziehen

- das bereichsübergreifende Denken mit der Zielrichtung auf integrierte Lösungen und die Zusammenarbeit in hochschulweit zusammengesetzten Arbeits- bzw. Projektgruppen sind bislang für die meisten Beteiligten ungewohnt

- der konkrete Nutzen des Gesamtprojektes ist für einige Entscheider in den Fakultäten und auch für einen Teil der direkt betroffenen IuK-Mitarbeiter noch nicht greifbar, wodurch ihre aktive Beteiligung in der Entwicklungsphase sowie die Akzeptanz in der Umsetzungsphase und somit die nachhaltige Wirkung der Veränderungen gefährdet sind
- bei dem ITS-Projekt handelt es sich um ein Innovationsprojekt, bei dem es gleichzeitig um umfassende Strukturveränderungen geht, weshalb die Beteiligten in der Loslösung von ihrem „bisherigen Selbstverständnis, von Strategien und Routinen des Alltagsgeschäfts“ (Heitger u a. 2002, S. 4) unterstützt werden müssen.

Daraus ergeben sich als Handlungsfelder für das Change-Management im ITS-Projekt: Es ist notwendig,

- ein Bewusstsein und Verständnis für den Changeprozess und für die Bedeutung der Partizipation aller Betroffenen zu schaffen (Change-Management-Kompetenz).
- eine hohe aktive Beteiligung bei den Betroffenen zu erreichen (Partizipation).
- den Nutzen der neuen Arbeitsweisen (Kooperation) und Denkweisen (Integration) sowie der technischen Projektinhalte für die Einzelnen aufzudecken.
- die Betroffenen in ihrem individuellen Veränderungsprozess zu unterstützen (Person).
- das Commitment und den Vorbildcharakter der Leitungsebene sowie deren Aktivitäten zur Unterstützung der Projektbeteiligten sichtbar zu machen.

### *2.3.2 Change-Architektur*

Als nächster Schritt wird, ausgehend von den aufgezeigten Handlungsfeldern, die grobe Struktur des Changevorhabens, d.h. seine tragenden Elemente und die zeitliche Abfolge der unterstützenden Maßnahmen des Veränderungsprozesses, in einer Change-Architektur erfasst (vgl. Königswieser u a. 2004, S. 54ff und S. 60ff). Diese Architektur stellt einen Rahmen dar, der die Interaktionsprozesse im Laufe des Changeprojektes vorstrukturiert, um den Beteiligten eine gewisse Sicherheit zu geben, jedoch auch für notwendige Prozessveränderungen offen ist und entsprechend der aktuellen Situation flexibel angepasst werden kann (vgl. Heitger u a. 2002, S. 19). Im ITS-Projekt wurde folgende Vorgehensweise gewählt:

- Zunächst wurden u a. mit Lenkungsrat, Projekt-, Arbeits- und Servicegruppen (s. o. Abb. 2) die relevanten personenbezogenen Elemente der Change-Architektur identifiziert und Heterogenität (unterschiedliche Fachkenntnisse

und Tätigkeitsbereiche, verschiedene Nutzerkreise) als Prinzip zur Zusammensetzung der Gruppen definiert.

- Davon ausgehend wurde ein Maßnahmenplan mit folgenden Elementen entworfen: Informationsveranstaltungen, Befragungen, Sitzungen, Social Events, Workshops, Trainings, Coachings, virtuelle Plattformen etc.
- Danach wurde die Architektur an Meilensteinen des Gesamtprojektes bzw. der einzelnen Teilprojekte (z. B. Projektstart, Zwischenevaluation) sowie an kulturellen, technischen oder strukturellen Spezifika der Organisation bzw. einzelner Abteilungen oder Teilprojekte (z. B. Planung von zusätzlichen Maßnahmen für einzelne Bereiche) ausgerichtet.

### 2.3.3 *Themen, Designs und Tools*

Im dritten Schritt werden die einzelnen Maßnahmen durch (1) die ausführliche Erfassung von Themenbereichen, (2) die Festlegung von konkreten Designs<sup>2</sup> und (3) die Auswahl geeigneter Tools (z. B. Modelle, Analyse- oder Erklärungskonzepte, Techniken) im Detail geplant. Im ITS-Projekt wurde beispielsweise, als eine konkrete Maßnahme des Handlungsfeldes Kooperation, für die Arbeitsgruppen-Leiter eine Reihe von drei aufeinander aufbauenden Workshops durchgeführt, um die Zusammenarbeit in den Arbeitsgruppen zu verbessern:

- Als relevante Themenbereiche wurden für diese Workshopreihe u. a. identifiziert: Kommunikation, Konfliktlösung, Leitung und Moderation von Gruppen, Reflexion von Arbeits- und Kooperationskulturen, Rollenklärung, bereichsübergreifende Information und Transparenz, Wertschätzung von Perspektivenvielfalt, Motivation zur und Nutzen von Zusammenarbeit.
- Der Ablauf der drei ganztägigen Workshops war in folgende inhaltliche Blöcke gegliedert: Einführung zum Thema, gemeinsame Vision und Ziele des Teilprojektes, gegenwärtige Qualität der Zusammenarbeit in und zwischen den Arbeitsgruppen (AGs), Ideal-Vorstellung der Zusammenarbeit in den AGs, Rollenklärung für die AG-Leiter und den Projektleiter, Vereinbarung von Regeln der Zusammenarbeit und Konfliktmanagement.
- Als Tools kamen dabei beispielsweise zum Einsatz: gemeinsame IST-Analyse in moderierter Plenumsdiskussion, lösungsorientierte Erarbeitung des Idealzustands in rotierenden Kleingruppen, Input zum Thema Rollenklärung, Coaching konkreter Konfliktfälle aus den AGs und moderierte Erarbeitung von Lö-

---

<sup>2</sup> Design meint hier: „die Gestaltung und Strukturierung eines bestimmten Prozessschrittes in der Beratung, eines einzelnen Architekturelementes“ (Königswieser u. a. 2004, S. 67).

sungsmöglichkeiten in der Gruppe der AG-Leiter als Form der kollegialen Beratung, Fokusanalyse zur Bestimmung der Konfliktebene und möglicher Ansatzpunkte für die Konfliktlösung (Personen-, Beziehungs-, Kontext- oder Sachebene).

#### *2.3.4 Evaluation und Anpassung der Planung*

Begleitend zur Umsetzung der einzelnen Maßnahmen wird parallel eine Evaluation ihrer Wirkungen durchgeführt und in der Folge zusätzlicher bzw. veränderter Bedarf in die Maßnahmenplanung einbezogen (vgl. Königswieser u a. 2004, S. 46f). Über den gesamten Veränderungsprozess hinweg wird somit die Change-Architektur immer wieder flexibel der sich durch die Maßnahmen verändernden Situation in der Organisation angepasst.

Bezogen auf das Beispiel der Workshopreihe für die ITS-Arbeitsgruppen-Leiter zeigte die Evaluation Bedarf für eine vertiefte Behandlung des Themas *Moderation von Gruppen*, woraufhin als ergänzende Maßnahme eine Trainingsreihe zu Moderationstechniken für die AG-Leiter etabliert wurde. Des Weiteren zeigte die Evaluation der Veranstaltungsreihe ITS-Forum beispielsweise, dass sie – durch die regelmäßigen, strukturierten und moderierten Treffen der ca. 25 Kernprojektmitarbeiter – erfolgreich zur Etablierung des ITS als Kommunikations- und Kooperationsraum beitragen konnte und dementsprechend weitergeführt werden sollte. Als zusätzliche Maßnahme soll in diesem Zusammenhang die Einrichtung einer virtuellen Plattform zur Abstimmung und zum Austausch der Projektbeteiligten angestoßen werden. Als weitere Maßnahme zur Einbindung von Universitätsangehörigen sollen sogenannte Expertenkreise berufen werden, die eine bereichsübergreifende, themenbezogene Kooperation von hochschulinternen Experten ermöglichen und die langfristige Verankerung des ITS-Kooperationskonzepts in der Universität gewährleisten sollen.

#### **2.4 Change-Management als Widerspruchsmanagement**

Literatur und Erfahrungen im ITS-Projekt zeigen, dass bei der Steuerung, d.h. bei der Planung und der Anpassung geplanter Maßnahmen, dieses komplexen – und in diesem Sinne eigentlich unsteuerbaren – Change-Vorhabens zahlreiche Widersprüche gemanagt werden müssen, worauf der nachfolgende Abschnitt näher eingehen wird.

#### 2.4.1 *Widerspruch 1: Verändern – Bewahren*

Ein gelungenes Changevorhaben zeichnet sich nicht dadurch aus, dass möglichst viele oder umfassende Veränderungen angestoßen werden, sondern vielmehr dadurch, dass eine Balance zwischen Verändern und Bewahren gehalten wird. In einem Aushandlungsprozess muss zwischen allen Beteiligten geklärt werden, ob und warum eine Notwendigkeit für Veränderung gesehen wird bzw. welche möglichen Konsequenzen es haben könnte, eine Veränderung durchzuführen oder sie zu unterlassen. Dies ist insbesondere wichtig, da Veränderungen jeder Art bei den betroffenen Mitarbeitern und Nutzern unterschiedliche Gefühle hervorrufen können, wobei sich zwei Gruppen als besonders wichtig für die Steuerung des Change-Prozesses erweisen: Die „Bewahrer des Gegenwärtigen“ zeichnen sich dadurch aus, dass die geplanten Veränderungen bei ihnen Vorsicht, Unsicherheit und Widerstand auslösen, während bei den „Motoren der Veränderung“ ein Klima des Aufbruchs, der Motivation und des aktiven Veränderungswillens vorherrscht. In den Arbeits- und Projektgruppen des ITS zeigte sich, dass „Bewahrer“ häufig als blockierend und bremsend für den Fortschritt der Diskussionen wahrgenommen werden, obwohl sie oft über gute, sachliche Gründe für ihren Wunsch nach einer Reduktion des Veränderungstempos verfügen (z. B. Vorteile von bestehenden Lösungen). Vertreter der Gruppe der „Motoren“ werden als vorstürmend und mitreißend empfunden, was sie zur Antriebsenergie von Veränderungen machen, aber auch zur Folge haben kann, dass noch nicht vollständig durchdachte Ideen beschlossen oder die Bedeutung von Veränderungen für die betroffenen Personen nicht berücksichtigt werden. Im ITS-Projekt hat sich im Umgang mit dem beschriebenen Widerspruch ein differenziertes Informationsmanagement als wesentlicher Erfolgsfaktor erwiesen: Da im IuK-Bereich derzeit an vielen Hebeln gleichzeitig angesetzt wird, muss den einzelnen Betroffenen von zentraler Stelle der Überblick und die Sicherheit darüber gegeben werden, was sich konkret für ihre Situation verändert und was eben nicht. Zudem stellte sich als wichtig heraus, dass die Projektbeteiligten die beschriebenen Verhaltensmuster und die dahinterliegenden Beweggründe kennenlernen und verstehen, um konstruktiv damit umgehen zu können. Des Weiteren benötigen beide Gruppen im Veränderungsprozess gleichermaßen Aufmerksamkeit durch den Change-Manager: Beispielsweise werden durch gezielte Moderation von Sitzungen die „Motoren“ in ihrem Veränderungstempo hin und wieder gedrosselt, um den Standort aller im Changeprozess zu bestimmen und sich gemeinsam denjenigen zuzuwenden, die dabei unbeachtet geblieben sind oder aktiven Widerstand gegen die Veränderung leisten.

#### *2.4.2 Widerspruch 2: Top down Steuerung – Bottom up Initiative*

Ein weiterer Widerspruch liegt darin, dass einige Aktivitäten in diesem Veränderungsprozess von der Führungsebene (top-down) gesteuert und andere von der Ebene der Betroffenen (bottom-up) initialisiert werden müssen. Heiger (2002) betont, dass es bei der Steuerung solcher Projekte darauf ankommt, dass die Führungsebene eine Gesamtorientierung gibt, zentrale Grundsatzentscheidungen trifft und Anreize für Projektaktivitäten setzt, während die Betroffenen Eigeninitiative aufbringen müssen, indem sie Innovationen initiieren und gewohnte Routinen aufgeben (vgl. Heitger u a. 2002, S. 40). Der Change-Manager kann dazu beitragen, in der Führungsebene das Bewusstsein zu schaffen, welche Aktivitäten von ihr top-down gesteuert werden müssen und wo es wichtig ist, Freiraum für bottom-up-Initiativen einzuräumen. Die Betroffenen kann er darin unterstützen, diesen Freiraum wahrzunehmen und die Maßnahmen entsprechend ihres Bedarfes zu gestalten. Im ITS-Projekt hat sich gezeigt, dass für die Beteiligten als Signale aus der Leitungsebene deren Commitment, eine gemeinsame Zukunftsorientierung (Vision), klare Veränderungsziele und Regeln für die Zusammenarbeit sowie ein deutlich umgrenzter Freiraum für die Mitgestaltung ihres zukünftigen Alltags erforderlich sind, um zur aktiven Mitarbeit am Changevorhaben zu motivieren und Akzeptanz für die Veränderungsmaßnahmen zu erhalten.

#### *2.4.3 Widerspruch 3: Innenorientierung – Außenorientierung*

Als wichtiges Thema hat sich auch die Frage nach der Balance zwischen einer Innen- und Außenorientierung im ITS-Projekt erwiesen (vgl. auch Heitger u a. 2002, S. 21). Als Außenorientierung ist zu verstehen, wenn der Fokus auf Nutzen der IuK-Anwender, auf Veränderungsprojekten anderer Universitäten oder Veränderungen im Hochschulumfeld liegt und damit Ziele wie Steigerung von Qualität und Quantität des Angebots oder wie Innovativität und Wettbewerbsfähigkeit der IuK-Leistungen in den Mittelpunkt rücken. War das zu Beginn der Planungsphase im ITS der Fall, richtete sich die Aufmerksamkeit mit der weiteren Formulierung des Projektantrags und mit Projektbeginn nach Innen und zwar auf die Gestaltung des Alltags der IuK-Anbieter, d.h. auf technische Lösungen sowie auf die Umstrukturierung von Arbeitsabläufen und deren Konsequenzen für die IuK-Mitarbeiter. Im Laufe des ersten Projektjahres führte ein steuernder Eingriff des Change-Managers wieder zu einer stärkeren Beachtung der Außenperspektive, woraufhin der Nutzen des Projektes für die Anwender klarer herausgearbeitet wurde und sie selbst noch stärker in den Veränderungsprozess einbezogen wurden, um diese wertvolle Quelle zu nutzen und eine erfolgreiche Umsetzung zu erreichen.

#### *2.4.4 Widerspruch 4: Ergebnisorientierung – Prozessorientierung*

Des Weiteren zeigt sich ein Widerspruch zwischen Ergebnis- und Prozessorientierung im ITS-Projekt. Aufgrund der begrenzten Ressourcen befinden sich die Projektbeteiligten im Zwiespalt: Sollen sie sich eher auf die Entwicklung von sichtbaren Ergebnissen in Form von neuen Infrastrukturen und Diensten konzentrieren oder sollen sie mehr den langfristig angelegten Struktur- und Kooperationsbildungsprozess in den Fokus nehmen. Beides ist zum Teil sogar miteinander verbunden: konkrete Ergebnisse hängen von einer funktionierenden Zusammenarbeit ab und diese wiederum setzt bestimmte Ergebnisse aus dem Projekt voraus. Dazu kommt weiterhin, dass sich durch die Strukturentwicklung die Ziele und Inhalte des Projektes selbst verändern, d.h. durch die kooperativen Strukturen entstehen zum Teil völlig neue, integriertere und innovativere Ideen als sie mit den alten Strukturen, Abläufen und Denkweisen in der Planungsphase möglich gewesen sind. Die Erfahrung zeigt, dass auch hier der Change-Manager aus seiner übergeordneten Perspektive hin und wieder steuernd eingreifen muss, um die Balance zwischen einer Konzentration auf den Veränderungsprozess und auf die inhaltlichen Ergebnisse zu erhalten bzw. wieder herzustellen.

#### *2.4.5 Widerspruch 5: Innovation – Tagesgeschäft*

Eine weitere Herausforderung ist die Aufgabe, Innovationen im Rahmen des ITS-Projektes zu schaffen, beispielsweise indem neue Angebote entwickelt, bestehende Konzepte auf neue Bereiche übertragen oder Abläufe auf eine neue Art und Weise durchgeführt werden. Der Widerspruch ergibt sich hier aus der Forderung an IuK-Mitarbeiter – aus ihrem Alltagsgeschäft heraus – innovative F&E zu betreiben bzw. an die neu eingestellten Projektmitarbeiter – ohne die Kenntnis des Alltagsgeschäftes – sinnvolle F&E zu leisten. Der Fokus der fest angestellten Mitarbeiter der IuK-Anbieter liegt in erster Linie auf dem zuverlässigen Betrieb der aktuellen Systeme und dem Angebot der gegenwärtig verfügbaren Dienstleistungen. Daher wirft die Entwicklung innovativer Angebote, die „anders“ als bisher sind, zwei Schwierigkeiten für sie auf: (1) die Innovationen können teilweise oder ganz gegen ihre bisherigen Überzeugungen und Strategien stehen; (2) sie müssen sich für die Beschäftigung mit innovativen Ideen eine Auszeit aus ihrem Tagesgeschäft nehmen, die aufgrund der knappen Ressourcensituation häufig nur schwer realisierbar ist. Dagegen können sich die neu eingestellten Projektmitarbeiter eher unvoreingenommen und mit dem notwendigen Freiraum auf innovative Ideen konzentrieren. Allerdings müssen sie sich dazu teilweise erst in das Alltagsgeschäft einarbeiten, um die für eine sinnvolle Veränderung notwendigen Zusammenhänge kennen zu ler-



nen. Trotzdem sind sie in vielen Fällen weiterhin von dem fachlichen Input und der Expertenmeinung der erfahrenen IuK-Mitarbeiter in den Einrichtungen abhängig. Die Herausforderung ist daher, im ITS eine geeignete Form der Zusammenarbeit zu etablieren und die Institutionalisierung eines regulären Innovationsprozesses im Rahmen einer universitätsweiten Innovationsstrategie zu erreichen, bei der IuK- und Projektmitarbeiter sowie Wissenschaftler und Studierende in enger Kooperation Innovationen entwickeln, die im Hochschulalltag zum Nutzen der gesamten Universität eingesetzt werden können (vgl. Sporer u a. 2007). Die Erfahrung zeigt, dass sich insbesondere die Gestaltung der Rahmenbedingungen für die Zusammenarbeit positiv auswirken kann: Durch eine ungewohnte zeitlich-räumliche Gestaltung (z. B. halbtägiger Workshop außerhalb der „Reichweite“ des Tagesgeschäfts statt der üblichen 2-stündigen Besprechungen am Arbeitsplatz mit häufigen Störungen) und durch die Anwendung von geeigneten Arbeitsmethoden (z. B. themenorientierte Gruppenarbeit, moderierte Themenarbeit, Kreativitätsmethoden, Fallarbeit, Rollenspiele etc. statt der gewöhnlichen Diskussionsrunden) kann ein Klima geschaffen werden, das neue Formen der Zusammenarbeit und innovatives Denken – losgelöst von der alltäglichen Arbeit – fördert. In den Workshops sollten die fest angestellten IuK-Mitarbeiter durch den Change-Manager zudem darin unterstützt werden, sich von ihrem bisherigen Selbstverständnis und den Strategien und Routinen ihres Alltagsgeschäftes zu lösen und die Chance, ihren zukünftigen Arbeitsbereich mitzugestalten, aktiv zu nutzen.

#### *2.4.6 Widerspruch 6: Abwarten – Entscheiden*

Letztlich ergibt sich aus der Tatsache, dass im ITS-Projekt an vielen Hebeln gleichzeitig gedreht wird, die Herausforderung, den richtigen Zeitpunkt für Entscheidungen und Handlungen zu identifizieren. Der Widerspruch liegt dabei darin, dass es – aufgrund vielfältiger Vernetzungen und Abhängigkeiten zwischen den Teilprojekten – für die Teilprojektleiter auf der einen Seite häufig vernünftig erscheint, mit der Definition eines abschließenden Anforderungskatalogs oder mit der eigenen Entscheidung für bzw. gegen ein System zu warten, bis entsprechende Informationen aus den anderen Teilprojekten vorliegen und im Rahmen einer Gesamtstrategie sinnvoll entschieden werden kann. Auf der anderen Seite hemmt dieses langwierige Abstimmungsverfahren und die Einbeziehung vieler Variablen und Anforderungen letztlich die Entscheidung in allen Teilprojekten und somit den Fortschritt des gesamten Projektes. Als Erfolgsfaktor erweist sich hierbei im ITS-Projekt die Erkenntnis, dass bei einem solch komplexen und vernetzten Projekt längere Entscheidungswege aufgrund der Vielzahl an zu berücksichtigenden Variablen und Personen, einerseits normal und von vornherein einzuplanen sind, jedoch während der

Projektarbeit auch ein gangbarer Mittelweg zwischen einem gründlichen Abstimmungsprozess und einer zügigen Entscheidung in den einzelnen Teilprojekten gefunden werden muss. Die Teilprojekte sollten dazu akzeptieren, dass Entscheidungen letztlich unter der Unsicherheit zu treffen, dass ein Teil der Anforderungen oder Rahmenbedingungen noch unklar ist, und dass bei einem solch komplexen Changevorhaben nicht alle Faktoren in eine Entscheidung einbezogen werden können. Der Change-Manager kann hierbei unterstützen, indem er den Abstimmungsprozess mit Blick auf einen sinnvollen Mittelweg moderiert, im Projektfortschritt neu entstehende Abhängigkeiten aufzeigt und über anstehende bzw. gefallene Entscheidungen in den einzelnen Teilprojekten sowie in den themenrelevanten Bereichen der Universität informiert.

### **3 Fazit und Ausblick**

Wirksam, schnell und flexibel Veränderungen gestalten zu können, wird als Wettbewerbsvorteil in Zukunft auch im universitären Bereich eine größere Rolle spielen. Die Erfahrungen im Projekt zum „Aufbau eines IT-Servicezentrums“ an der Universität Augsburg zeigen, dass zur Begleitung solcher komplexen und dynamischen Veränderungsprozesse professionelles Change-Management erforderlich ist und in der Universität langfristig eine entsprechende Change-Kompetenz aufgebaut werden muss.

Die Maßnahmen des Change-Managements im ITS-Projekt orientieren sich bereits seit Beginn an Umsetzung und Nachhaltigkeit der geplanten Veränderungen. Dabei werden durch gezielte Partizipationsmaßnahmen „die vom Veränderungsprozess Betroffenen“ zu „am Veränderungsprozess Beteiligten“ gemacht, um diese wertvollen Ressourcen für das Changevorhaben zu nutzen und die Akzeptanz der Veränderungen in der Universität zu verbessern. Neben den inhaltlichen Fortschritten, zeigt sich als erste Erfolge, dass die Aktivitäten der Teilprojekte und die Etablierung der neuen Strukturen sowie die Einbeziehung der Nutzer in vielen Bereichen bereits zusätzliche, über die ursprüngliche Planung hinausgehende Kooperationen und neue Projekte angestoßen haben. Zudem arbeiten Universitätsangehörige aus den verschiedensten Einrichtungen und Fakultäten im Kooperationsraum ITS motiviert zusammen und gestalten integrierte Konzepte, u a. deshalb, weil sie gezielt im Wandel ihrer Arbeitsweisen und Denkmuster unterstützt werden. Jedoch wird auch klar, dass ein solch komplexer Veränderungsprozess Zeit benötigt und dass – wenn man sich vorgenommen hat, dass alle Beteiligten gemeinsam am Ziel ankommen – noch einige Herausforderungen zu bestehen sind.

Als weitere Schritte stehen für das Change-Management die Begleitung der Teilprojekte bei der Entwicklung eines bereichsübergreifenden Betreuungs- und Beratungskonzepts für den IuK-Bereich, bei der Einrichtung des ITS-Kundenbüros und bei der Umgestaltung der Dienstleistungserbringung in ITS-Servicegruppen im Mittelpunkt. Zudem gilt es, eine tragfähige Lösung zu finden, wie durch entsprechende Maßnahmen und über elektronische Plattformen sowohl die große Zahl der Universitätsangehörigen als auch die direkt am Projekt Beteiligten jederzeit über den Projektstatus und die erzielten Fortschritte informiert sein können. Nicht zuletzt steht die Erkenntnis, dass das ITS-Projekt nicht als einmaliger Veränderungsprozess zu sehen ist, sondern dass der Wandel im IuK-Bereich weiter gehen wird und eine entsprechende Change-Strategie dafür zu etablieren ist.

## **4 Literaturangaben**

- Doppler, K.; Fuhrmann, H.; Lebbe-Waschke, B.; Voigt, B. (2002): Unternehmenswandel gegen Widerstände. Change Management mit den Menschen. Frankfurt/Main: Campus Verlag GmbH.
- Doujak, A.; Endres, T.; Schubert, H. (2004): IT & Change mit Wirkung. In: Boos, F.; Heitger, B. (Hrsg.): Veränderung – systemisch. Management des Wandels – Praxis, Konzepte und Zukunft. Stuttgart: Klett-Cotta Verlag, S. 313-317.
- Eichner, L.; Hohoff, U. (2006): Der Aufbau eines IT-Servicezentrums an der Universität Augsburg. In: Degkwitz A. / Schirmbacher, P.: Informationsinfrastrukturen im Wandel. Informationsmanagement an deutschen Hochschulen. Preprint Version, S. 21-35.  
[http://www.dini.de/documents/DINI\\_Informationsinfrastrukturen\\_Preprint.pdf](http://www.dini.de/documents/DINI_Informationsinfrastrukturen_Preprint.pdf)
- Heitger, B.; Doujak, A. (2002): Harte Schnitte, neues Wachstum: Die Logik der Gefühle und die Macht der Zahlen im Changemanagement. Frankfurt u. a.: Redline Wirtschaft bei Ueberreuter.
- Königswieser, R.; Hillebrand, M. (2004): Einführung in die systemische Organisationsberatung. Heidelberg: Carl-Auer-Systeme Verlag.
- Königswieser, R.; Exner, A. (1997): Systemische Intervention. Architekturen und Designs für Berater und Veränderungsmanager. Stuttgart: Klett-Cotta.
- Sporer, T.; Fahrner, U.; Mauermann, L. (2007): Gestaltung von digitalen Lehr-Lern-Szenarien durch offene Innovationsstrategien an Hochschulen. In: A. Oßwald, M. Stempfhuber & C. Wolff (Eds.), Open Innovation. Proc. d. 10. Internationalen Symposiums für Informationswissenschaft, Köln, Mai 2007 (in diesem Band). Konstanz: UVK.



# Dokumentvorlagen zur Erstellung digitaler Hochschulschriften mit LaTeX\*

## Ergebnisse des DFG-Projektes DissOnline Tutor

*Verena Gondek*

Universität Duisburg-Essen, Fachbereich Mathematik  
Forsthausweg 2, 47057 Duisburg  
*verena.gondek@uni-due.de*

### **Zusammenfassung**

Im DFG-Projekt DissOnline Tutor wurden für Autoren von wissenschaftlichen Hochschulschriften einfach handhabbare Dokumentvorlagen für Textverarbeitungssysteme entwickelt. Auf diese Weise werden die Autoren beim Erstellen ihrer Publikationen technisch unterstützt sowie deren Langzeitarchivierbarkeit durch Einhalten erforderlicher Standards verbessert.

### **Abstract**

In the DFG project DissOnline Tutor easy to use document samples for authors of scientific publications for text processing systems develops were developed. Authors are technically supported with providing their publications as well as longterm archiving is improved by keeping necessary standards.

## **I DFG Projekt DissOnline Tutor**

Eine wissenschaftliche Arbeit elektronisch zu publizieren stellt viele Autoren vor nahezu unüberwindbare Hürden. Zwar ist der Umgang mit Textverarbeitungssoftware üblich, jedoch nur mit dem Ziel, einen Ausdruck der Arbeit auf Papier zu erzeugen. Wissen über den effektiven Umgang mit solchen Systemen für das elektronische Publizieren ist in den wenigsten Fällen vorhanden, denn dabei sind unter anderem Gesichtspunkte wie standardisierte Dokumentformate, Metadaten und

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 465-468.

Langzeitarchivierung in einer Bibliothek von enormer Bedeutung. Diese Aspekte müssen vom Autor schon bei der Generierung seines Dokuments berücksichtigt werden, da es im Nachhinein kaum realisierbar ist, eine fertige Arbeit noch an entsprechende Standards anzupassen. Die einzige Ausnahme bildet in eingeschränktem Rahmen LaTeX, sofern auch die Quelldateien und nicht nur ein fertiges pdf-/dvi-Dokument vorliegen, denn hier kann man auch im Nachhinein noch relativ einfach Formatierungen vornehmen.

Im Hinblick auf die Problematik schon beim Schreiben einer wissenschaftlichen Arbeit ohne großen zusätzlichen Aufwand vorgegebene Standards zu erfüllen, wurden im Rahmen des DFG-Projektes DissOnline Tutor an der Universität Duisburg-Essen sowie der Humboldt-Universität Berlin in Zusammenarbeit mit der Deutschen Nationalbibliothek Dokumentvorlagen für LaTeX (Duisburg) und WYSIWYG-Systeme (Berlin) entwickelt. Damit werden den Autoren einfache Hilfsmittel zur Verfügung gestellt, die sich an gängige Schreibumgebungen anpassen. Dennoch müssen die Autoren den Umgang mit diesen Systemen beherrschen. Deshalb nimmt DissOnline Tutor seinen Namen ernst und bietet auf *DissOnline.de* Hilfe in Form von Anleitungen, Beispielen und einer Frage-Antwort-Sammlung für den Umgang mit diesen Tools an.

## 2 LaTeX Dokumentvorlage

Die Dokumentvorlage für LaTeX setzt sich aus fünf einzelnen Files zusammen. In der Hauptdatei, werden die übrigen Dateien eingebunden und der Autor gliedert alle weiteren Files, die er für seine Arbeit benötigt oder erstellt (Kapitel der Arbeit, gewünschte Style Files etc.) ebenfalls an dieser Stelle in das Dokument ein. In der zugehörigen Dokumentklasse DissOnlineLatex werden alle erforderlichen Pflichtpakete zur Nutzung der Vorlage mit einigen voreingestellten Paketoptionen eingebunden. Der Autor hat jedoch die Möglichkeit in der Hauptdatei selbständig weitere oder auch weniger Optionen zu wählen. Darüber hinaus werden in der Dokumentklasse grundlegende graphische Formatierungen festgelegt. Diese sind aber vor allem als Beispiel zu verstehen, denn in Bezug auf das Layout seines Dokumentes hat der Autor relativ große gestalterische Freiheit. Die wichtigste Aufgabe der DissOnlineLatex Klasse ist es jedoch Warnmeldungen zu generieren, die bei der Kompilation angezeigt werden, falls der Autor Konventionen bei der Erstellung des Dokumentes nicht berücksichtigt haben sollte. Hat er beispielsweise ein Style-File verwendet, dass entweder nicht kompatibel mit der Dokumentvorlage ist oder bei dessen Verwendung andere Schwierigkeiten auftreten können weil es veraltet ist

oder Ähnliches, so erhält der Autor beim Kompilieren eine entsprechende Warnung und gleichzeitig einen Hinweis, welches Paket er alternativ verwenden könnte.

Über das dritte File der Dokumentvorlage (*metadata.tex*) werden alle benötigten Metadaten des Autors eingelesen. Er trägt diese einmalig in die entsprechend gekennzeichneten Felder ein. Von dort aus werden sie zur weiteren Verarbeitung in die Dokumentklasse eingelesen. Auf diese Weise stehen sie auch bereits zur Generierung einer Titelseite für das Dokument zur Verfügung. Die vierte Datei der Dokumentvorlage (*titlepage.tex*) enthält einen Vorschlag für das Layout der Titelseite des zu erstellenden Dokuments. An diese Formatierung ist der Autor natürlich nicht gebunden, da dieser vor allem die Vorgaben seiner Universität, bzw. Universitätsbibliothek berücksichtigen muss. Die letzte mitgelieferte Datei (*abstract.tex*) muss den Abstract der Arbeit enthalten, sofern der Autor seiner Arbeit eine solche Zusammenfassung beifügen möchte. Dass eine Arbeit ein Abstract enthält ist nicht verpflichtend, höchstwahrscheinlich wird aber jede Bibliothek bei der Abgabe einer Arbeit auch eine Zusammenfassung zu Archivierungszwecken fordern.

Die entstandene Dokumentvorlage ist übersichtlich und weist eine einfache Dateistruktur auf. Dies ist ein entscheidender Vorteil gegenüber einigen anderen Vorlagen für LaTeX-Dokumente, welche häufig kompliziert und unübersichtlich sind. Auf diese Weise werden auch Benutzer angesprochen, die bisher nur Grundkenntnisse in LaTeX haben. Um den Umgang mit der Dokumentvorlage zusätzlich zu erleichtern, sind die vorhandenen Möglichkeiten, etwaige Schwierigkeiten etc. in einer Broschüre dokumentiert. Diese ist zusammen mit der Vorlage auf [DissOnline.de](http://DissOnline.de) erhältlich. Für WYSIWYG-Systeme wurde Ähnliches umgesetzt. Nutzer von Microsoft Word oder auch OpenOffice/StarOffice finden auf [DissOnline.de](http://DissOnline.de) ebenfalls sehr gute Dokumentvorlagen für wissenschaftliche Hochschulschriften und entsprechende Anleitungen.

### 3 Offene Formate und Archivierbarkeit

Welche Schreibumgebung ein Autor wählt ist häufig abhängig von seinem Fachgebiet. LaTeX ist vor allem im mathematisch-naturwissenschaftlichen Bereich weit verbreitet. Doch auch für Wissenschaftler anderer Fachgebiete kann es lohnend sein, sich außerhalb der Welt von Word und OpenOffice umzusehen, denn auch LaTeX-Quelltext ist relativ gut lesbar, vor allem, wenn er überwiegend Fließtext enthält. Dieser Aspekt hat einen entscheidenden Vorteil im Hinblick auf nachhaltige Archivierbarkeit von Dokumenten. Es ist unmöglich zu sagen, wie sich dieser

Bereich zukünftig entwickeln wird. Der Text in einer LaTeX-Quelldatei aber, kann mit jedem beliebigen reinen Texteditor gelesen werden. Möglicherweise sind manche Befehle in Formeln oder Ähnliches in Zukunft nicht mehr lesbar. Der Text an sich wird jedoch erhalten bleiben. Des Weiteren fordert DissOnlineLaTeX zum Verwenden aktueller Pakete auf, um aktuelle Standards einzuhalten. Durch die automatisierte Warnung vor nicht mehr unterstützten oder bereits veralteten Paketen, wird dem Autor ein nützliches Werkzeug an die Hand gegeben, um die Langzeitarchivierbarkeit seiner Publikationen zu verbessern.



# Die Initiative Digital Peer Publishing NRW – Erfahrungen mit dem kollaborativen Publikations- und Kommunikationsmodell\*

*Cordula Nötzelmann*

Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen

## **Zusammenfassung**

Dieser Beitrag stellt dar, unter welchen Voraussetzungen die Publikationsplattform Digital Peer Publishing NRW (DiPP NRW; <http://www.dipp.nrw.de>) nach Ablauf des Projektförderzeitraums operiert.

In einer kritischen Bestandsaufnahme geht der Beitrag auf die aktuellen organisatorischen Rahmenbedingungen der Initiative und ihre Kooperationsmodelle mit den beteiligten Institutionen ein. Er bietet weiterhin einen Überblick über den Arbeitsstand der bereits etablierten eJournals, wie auch den Status Quo bei inzwischen neu gewonnenen DiPP-Partnern und -Nutzern.

Neben dem eigens für DiPP NRW entwickelten Geschäftsmodell-Entwurf behandelt der Beitrag die nächsten Schritte zur Sicherung der Nachhaltigkeit und perspektivische Formen der Zusammenarbeit, z. B. mit dem auf medizinische Inhalte spezialisierten Content- und Journal-Portal German Medical Science (GMS; <http://www.egms.de/de/>).

Technische Weiterentwicklungen der DiPP-Plattform werden vor allem im Hinblick auf die für das wissenschaftliche Informationsmanagement relevanten Anpassungen im Bereich „Web 2.0“ sowohl retro- als auch prospektiv betrachtet. Die zur Zeit primär gewählte Publikationsform „eJournal“ eignet sich dabei besonders, um Tools zu unterstützen, die den wissenschaftlichen Diskurs weiter vereinfachen und transparent abbilden können. Der Beitrag verweist darüber hinaus auf weitere Publikationsmöglichkeiten, die die DiPP-Plattform bietet. Weiterführende Aspekte zur Einbindung des Projekts in die Dienstleistungen des Hochschulbibliothekszentrums NRW (hbz; [www.hbz-nrw.de](http://www.hbz-nrw.de)) im Bereich Publikationssysteme runden den Beitrag ab.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 469-478.

## I Bestandsaufnahme

„Digital Peer Publishing NRW“ ist bundesweit das erste Beispiel für die Verankerung von Open Access Publikationsstrukturen für eJournals in den Hochschulalltag eines Bundeslandes. Als Initiative des Landes Nordrhein-Westfalen wurde DiPP NRW in der Aufbau- und Konsolidierungsphase mit acht Zeitschriften an Universitäten und Fachhochschulen vom MIFWT im Zeitraum von 2004 bis 2006 gefördert.

Im Produktivbetrieb entwickelt das Hochschulbibliothekszenrum NRW (hbz) die Softwareplattform weiter und begleitet die Redaktionen unterstützend. Das Produkt hat sich in der deutschsprachigen Open Access-Szene positioniert und gilt als eines der ausgereiftesten Publikationssysteme für eJournals in Deutschland.<sup>1</sup>

### 1.1 Entwicklung der eJournals

Die laufenden Zeitschriften konnten selbst nach einem für das Publikationswesen äußerst eng bemessenen Zeitraum eine beachtliche Produktion von qualitätsgesicherten Publikationen mit stetig steigenden Zugriffszahlen und ersten Zitationen vorweisen und dies in begrenztem Rahmen auch ohne zusätzliche Fördermittel fortführen.

Nach dem Ende der Förderungsdauer steht DiPP auf einer soliden Basis, um weitere Partner und Projekte mit in den Produktionsbetrieb aufzunehmen. Die Plattform hat sich zum Jahreswechsel 2005/2006 für weitere Redaktionen geöffnet. In der Folge sind mehrere neue eJournal-Projekte zu verzeichnen, die sich derzeit kurz vor dem Start, im Teststadium oder in der konkreten Anbahnung befinden. Im März 2007 laufen dreizehn, zum Großteil international vernetzte eJournals auf der DiPP-Plattform, die ein breites Spektrum an wissenschaftlichen Disziplinen abdecken.

---

<sup>1</sup> Vgl. aus der Wissenschaft die Übersichtseite der Universität Bielefeld über *Nationale und Internationale Aktivitäten im Bereich Wissenschaftliches Publizieren*; URL: <http://www.uni-bielefeld.de/ub/wp/projekt.htm>, aus der Welt der Infrastrukturdienstleister die Informationsseite *Akteure und Projekte zum "elektronischen Publizieren" in der deutschen Bibliothekslandschaft* auf dem Bibliotheksportal des Kompetenznetzwerks für Bibliotheken. URL: <http://www.bibliothekportal.de/hauptmenue/themen/digitale-bibliothek/elektronisches-publizieren/akteure-und-projekte>

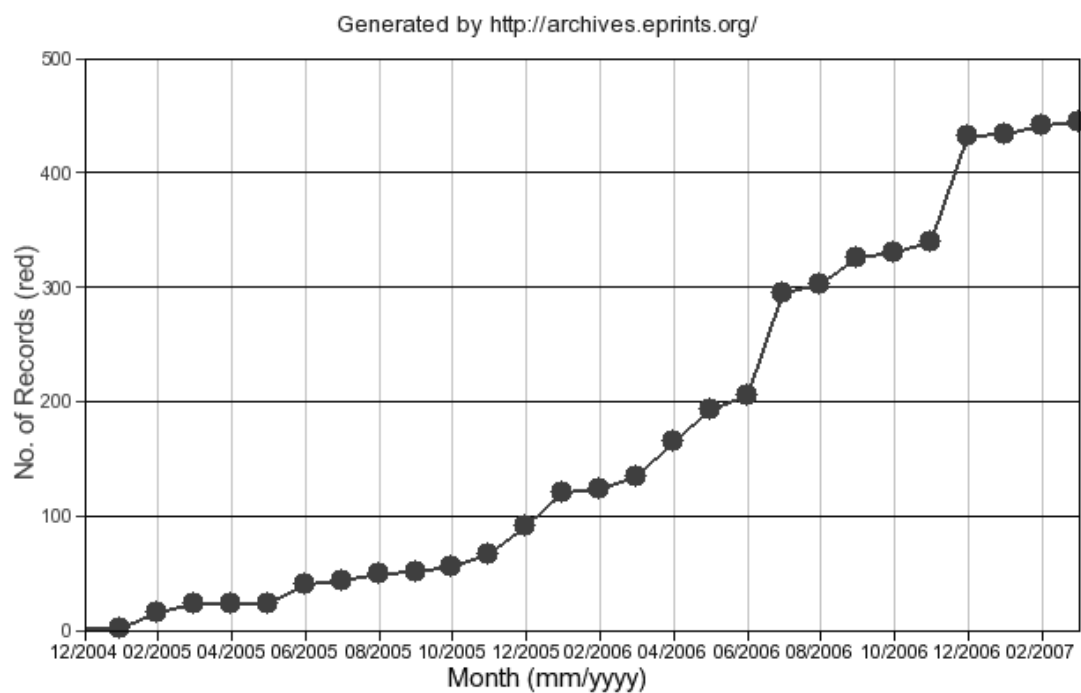


Abbildung 1: Anzahl der Publikationen in DiPP-Repository 2

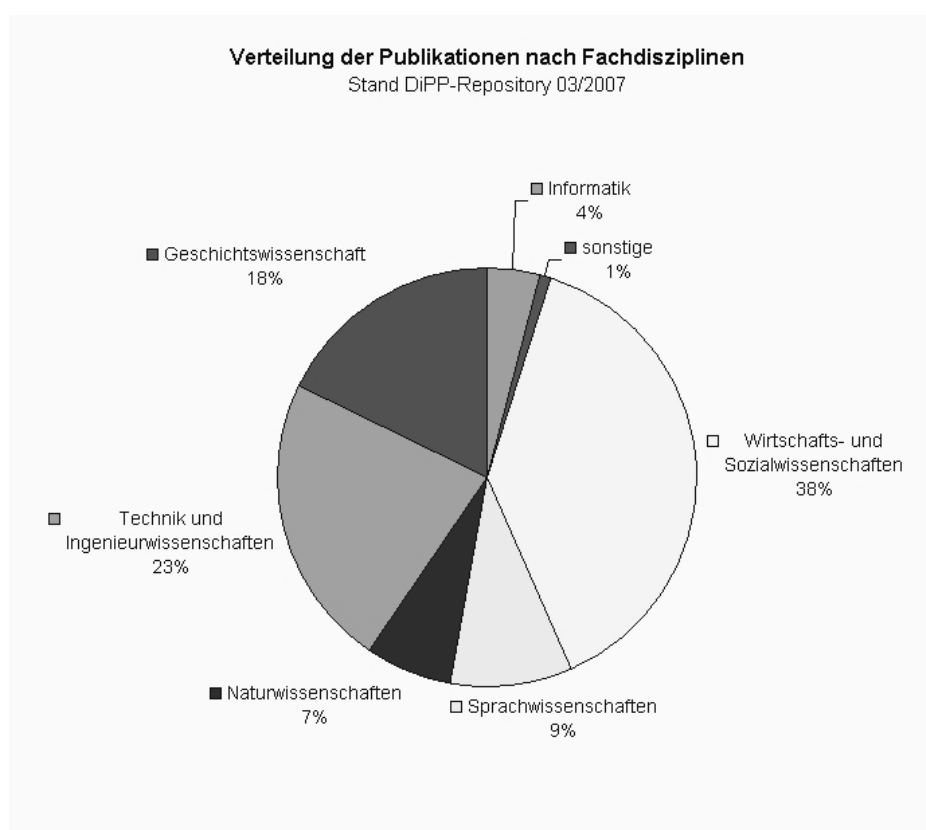


Abbildung 2: Verteilung der Publikationen nach Fachdisziplinen im DiPP-Repository. 3

<sup>2</sup> Die Angaben sind Näherungswerte, da Ausgaben der Zeilenblicke vor 2005 nicht berücksichtigt sind.

Die verschiedenen redaktionellen Kooperationsmodelle<sup>4</sup> sind mittlerweile erfolgreich erprobt und bewährt. Die konzeptionelle Nähe der DiPP-Initiative zu den wissenschaftlichen Bibliotheken ermöglicht es, redaktionelle Arbeiten, z. B. in Form von Redaktionstandems<sup>5</sup>, auf mehrere Partner zu verteilen.

Herausgeber- und Redaktionstreffen, Mailinglisten und ein gemeinsames Arbeits-Wiki fördern die Kommunikation der DiPP-Anwenderinstitutionen untereinander und mit dem hbz in Köln.

## 2 Nachhaltigkeit

Im folgenden wird beschrieben, in welcher Form DiPP organisatorisch in die Trägerschaft des hbz eingebunden wurde, um die Voraussetzungen für Innovationsentwicklungen und -transfer weiterhin zu gewährleisten. Auch technische Maßnahmen und Überlegungen zu einem Geschäftsmodell-Entwurf sind hier zu nennen.

### *Organisatorische Rahmenbedingungen*

Das mit der Weiterführung der Dienstleistung beauftragte hbz hat im Zuge der Bündelung seiner Services rund um Publikationssysteme die DiPP-Initiative in einen größeren Zusammenhang gestellt. Seit März 2007 werden unter neuer Projektleitung in der zeitgleich neu strukturierten Gruppe „Publikationssysteme“ die laufenden und im Aufbau befindlichen Publikationsservices des hbz organisatorisch und funktional zusammengeführt mit dem Ziel, auf diesem Gebiet Ressourcen zu vernetzen und Synergien zu erzielen.

Die technologische Basis des Systems, die sich im Produktionsbetrieb bereits bewährt hat, wird kontinuierlich aktualisiert und an neue Entwicklungen angepasst. Gleiches gilt für die Lizenz, unter der das elektronische Publizieren mit DiPP funktioniert: die für den deutschen Rechtsraum entwickelte, für digitale Dokumente zugeschnittene modulare Digital Peer Publishing Lizenz (DPPL) wird inzwischen auch von elektronischen Zeitschriften verwendet, die nicht auf der DiPP-Plattform

---

<sup>3</sup> S. Fn 2.

<sup>4</sup> Horstmann, Wolfram: „Kooperationsmodelle für Open Access eJournals in der Publikationsinitiative DiPP NRW“, in: *Knowledge eXtended (3. Konferenz der Zentralbibliothek), Schriften des Forschungszentrums Jülich, Reihe Bibliothek*, Band 14 (2005), S. 109-122.

<sup>5</sup> Nötzelmann, Cordula und Sören Lorenz: „Das Redaktionstandem als innovatives Kooperationsmodell zwischen Fachwissenschaftlern und Bibliothekaren am Beispiel des Open Access E-Journals *Brains, Minds & Media*“, ebd., S. 91-100.

laufen<sup>6</sup>. Die Lizenz ist z. B. mittlerweile auch als Kriterium für einen bundesweiten Standard im DINI-Zertifikat 2007 verankert.<sup>7</sup> Die DiPP-Lizenzen werden in Zusammenarbeit mit dem ifross<sup>8</sup> auf dem neuesten Stand gehalten.

## 2.2 Technische Weiterentwicklung

DiPP versteht sich langfristig als Inkubationszentrum für innovative, wissenschaftliche Publikationsformen. Die Plattformbetreiber reagieren auf neue Erfordernisse in den wissenschaftlichen Publikations- und Kommunikationsstrukturen. Möglichkeiten, das Produkt perspektivisch zu einer universellen Publikationsplattform auszubauen, die unter noch zu bestimmenden Voraussetzungen bestehende internationale Standards integrieren und forschungsaffine Kommunikationsstrukturen unterstützen kann, befinden sich derzeit in Prüfung.

### 2.2.1 Die Editorial Toolbox – Weiterführende Features auf Basis von Plone

Die Systemkomponenten Zope (Applikationsserver), Plone (Content Management System) und Fedora (Repository)<sup>9</sup> werden kontinuierlich auf den neuesten Stand gebracht und getestet. Dabei wird Sorge dafür getragen, die Migration der Journals auf neue Systemversionen peu a peu vorzunehmen. Neue DiPP-Partner starten selbstverständlich bereits auf dem jeweils aktuellen technischen Stand der Plattform.

Die zur Zeit primär gewählte Publikationsform „eJournal“ eignet sich dabei besonders, um Tools zu unterstützen, die den wissenschaftlichen Diskurs weiter vereinfachen und transparent abbilden können. Neue Funktionalitäten und Tools aus dem Kontext „Web 2.0“ sind über DiPP verfügbar: Neben RSS-Feeds, die bereits seit letztem Jahr nutzbar sind, können die Redaktionen ihre Zeitschriften um ein Blog, ein Wiki oder um ein Forum erweitern, die Artikel nicht nur mit URNs, sondern

---

<sup>6</sup> Dies tut zum Beispiel das elektronische *International Journal of Conflict and Violence*. URL: <http://www.ijcv.org/>

<sup>7</sup> Die Deutsche Initiative für Netzwerkinformation e.V. stellt mit ihrem Zertifikat, das grundlegende Standards zugrundelegt, die Qualität von Publikationsservices sicher. [http://www.dini.de/documents/dini\\_zertifikat\\_2007.pdf](http://www.dini.de/documents/dini_zertifikat_2007.pdf)

<sup>8</sup> Das Institut für Rechtsfragen der Freien und Open Source Software ist seit Projektbeginn Partner zur Entwicklung der DPPL. URL: <http://www.ifross.de/>

<sup>9</sup> Zur Architektur der DiPP-Plattform vgl. Horstmann, Wolfram und Jochen Schirrwagen, Peter Reimer: „Multi-level eJournal support structures in the initiative Digital Peer Publishing NRW.“ Joint Workshop on Electronic Publishing, Organised by Delos, SVEP and ScieCom. Lund University Libraries, Sweden, 2005. URL: [http://www.dipp.nrw.de/ueberdipp/publikationen/horstmann\\_lund\\_full.pdf](http://www.dipp.nrw.de/ueberdipp/publikationen/horstmann_lund_full.pdf)

bei Bedarf mit DOIs und OpenURLs versehen. Die auf Basis von Plone geschaffene Editorial Toolbox unterstützt den Redaktions-Workflow durch neue Möglichkeiten des Peer-Reviews (offenes und geschlossenes Peer-Review) sowie Nutzungs- und artikelbasierte Statistiken. Eine Kommentarfunktion befindet sich in Planung.

Im Zuge des Ausbaus der Plattform ist mittel- bis langfristig die Einbeziehung von externen Diensten geplant, etwa eines Print-On-Demand-Anbieters, oder die Aufnahme internationaler, offener Standards in das DiPP-Angebotsspektrum, die das wissenschaftliche Publizieren in eJournals unterstützen. Dies geschieht in enger Zusammenarbeit mit einzelnen Journalredaktionen.

### *2.2.2 Ausbau des DiPP-Repositorys für komplexe Objekttypen*

Besonderes Potenzial für weiterführende Publikationsdienstleistungen bietet die Repository-Software Fedora, deren System dank des Konzepts digitaler Objekte, denen beliebige Datenströme zugeordnet werden können, flexibel erweiterbar und auf konkrete Anwendungsfälle anpassbar ist. Fedora verfügt über ein Disseminator-Feature, das z. B. die Ergänzung zusätzlicher Metadaten oder die Weiterverarbeitung digitaler Objekte erlaubt.

Da es technisch keinen Unterschied macht, ob ein Repository für eJournals, Dissertationen, Preprints oder Postprints eingesetzt wird, eröffnet die Verwendung von Fedora auf der DiPP-Plattform neben der bisher vornehmlich gewählten eJournal-Variante weitere Publikationsmöglichkeiten. Denkbar wäre der Einsatz für Zwecke im Bereich der institutionellen Repositorien, vor allem bei der Implementierung anwendungsspezifischer Frameworks, mit dessen Hilfe verschiedene Nutzungsszenarien umgesetzt werden können, die über das bloße Einstellen und Verwalten von Inhalten hinausgehen.<sup>10</sup>

## **2.3 Geschäftsmodell**

Die Frage nach der nachhaltigen Finanzierung von Open Access-Angeboten ist sowohl im deutschsprachigen als auch im anglo-amerikanischen Raum weiterhin grundsätzlich noch nicht abschließend geklärt. Es haben sich inzwischen jedoch

---

<sup>10</sup> Beispiele für den Einsatz von Fedora als Repository-Lösung sind das Projekt *eSciDoc* der Max-Planck-Gesellschaft in Kooperation mit dem FIZ Karlsruhe, URL: <http://www.escidoc-project.de/de/materialien-downloads.html>, oder das auch für multimediale Inhalte genutzte institutionelle Repository der University of Queensland. URL: <http://espace.library.uq.edu.au/>.

einige mögliche Geschäftsmodelle formiert, wenn auch nur zum Teil etabliert<sup>11</sup>. Die derzeit erprobten Formen sind die Gründung einer Stiftung bzw. eines Vereins, die institutionelle Mitgliedschaft, oder, wie im Falle von German Medical Science, eine Beteiligung der Fachgesellschaften.<sup>12</sup> Wünschenswert wäre im Zusammenhang mit konzeptionellen Überlegungen zu Open Access-Geschäftsmodellen auch ein im wissenschaftlichen Alltag bereits etablierteres Instrumentarium zur Beantragung von Mitteln zur Deckung der Publikationskosten.<sup>13</sup>

Auch als Landesinitiative ist DiPP NRW in der Pflicht, neben den genannten organisatorischen, rechtlichen und technologischen Maßnahmen zur Sicherung des nachhaltigen DiPP-Angebots auch kontinuierlich förderungsunabhängige Finanzierungsmöglichkeiten zu prüfen. Seit Ende der Förderung wurde in regelmäßigen Abständen geprüft, ob die Voraussetzungen und der Zeitpunkt<sup>14</sup> gegeben sind, um ein Betriebsmodell auf finanzieller Basis zu entwickeln. Dies war bisher nicht der Fall. Vor dem neuen Hintergrund der Konsolidierung der hbz-Publikationsservices ist es Aufgabe der DiPP-Betreiber, verschiedene Konzepte zu prüfen und bisher entwickelte Ansätze zu überdenken.

Ideen zur Finanzierung der Journals kommen auch aus den Redaktionen, z. B. durch das Erwirtschaften von Werbeeinnahmen, etwa durch Nutzung des Google AdSense-Programms. Da solche Maßnahmen wegen der sehr heterogenen Nutzer-

---

<sup>11</sup> Vgl. Bauer, Bruno: „Kommerzielle Open Access Publishing-Geschäftsmodelle auf dem Prüfstand: ökonomische Zwischenbilanz der *Gold Road to Open Access* an drei österreichischen Universitäten.“ *GMS Med Bibl Inf* 2006;6(3):Doc32. URL:

<http://www.egms.de/en/journals/mbi/2007-6/mbi000050.shtml>

<sup>12</sup> Vgl. Neumann, Jan: „Auf dem Weg zu einem Open-Access-Geschäftsmodell“. *Open Source Jahrbuch 2006, Kapitel Wissenschaft<sup>2</sup>: Open Access*. URL:

[http://www.opensourcejahrbuch.de/download/jb2006/chapter\\_07/osjb2006-07-01-neumann](http://www.opensourcejahrbuch.de/download/jb2006/chapter_07/osjb2006-07-01-neumann)

<sup>13</sup> Vgl. *Wissensmanagement. Zeitschrift für Innovation*, 1/2006, Lemmens Verlag, Bonn. Dort heißt es unter der Überschrift: „Was ist zu tun?“ unter anderem:

„Eine wichtige Voraussetzung für Open Access ist die finanzielle Absicherung von Publikationskosten. Publikationskosten sind Forschungskosten und müssen daher im Budget jedes Forschungsvorhabens fest verankert sein. Insbesondere für die Finanzierung originärer Publikationen in renommierten Open-Access-Journals oder in wichtigen Hybridzeitschriften, die Wissenschaftlern die Möglichkeit bieten, eigentlich lizenzpflichtige Beiträge durch Zahlung einer Publikationsgebühr entgeltfrei zugänglich zu machen, müssen Mittel verfügbar sein.“ (S. 24) URL:

[http://www.lemmens.de/verlag/zeitschriften/wima\\_special\\_aktuell/special1\\_06.pdf](http://www.lemmens.de/verlag/zeitschriften/wima_special_aktuell/special1_06.pdf)

<sup>14</sup> Zu diesen Voraussetzungen gehört, etwa zur Erhebung von Autorengebühren bzw. Page-Charges, die entsprechende Reputation einer Zeitschrift, die sich erst im Laufe der Zeit entwickelt.

struktur und disziplinären Ausrichtung der beteiligten eJournals nicht unbesehen für alle DiPP-Journals gleichermaßen praktikabel sein können, soll ein innovatives Pilot-Journal diese Möglichkeit zunächst ausloten. Nach einer Evaluationsphase ist die Übertragbarkeit auf weitere geeignete DiPP-Anwender denkbar.

## 2.4 Kooperationen

Wie bisher deutlich wurde, arbeitet DiPP NRW mit den beteiligten Redaktionen bei der Implementierung neuer Services und Funktionen eng zusammen, doch ergeben sich auch Berührungspunkte mit Anwendern anderer Plattformen, die Komponenten der DiPP-Services zusätzlich nutzen möchten. Die Rede war bereits von Nutzern der DiPP-Lizenz. Zu nennen ist hier auch das Projekt *elanguage.net*<sup>15</sup>, das unter der Schirmherrschaft der Linguistic Society of America vom Lehrstuhl für Englische und Amerikanische Sprachwissenschaft an der Universität Düsseldorf betreut wird. Das hbz betreut das im Rahmen des Projekts entstehende *Matrixjournal* technisch auf der DiPP-Ebene und gewährleistet die dauerhafte Verfügbarkeit; projektspezifische Features fließen im Gegenzug in die DiPP-Plattform mit ein. Die bereits engen Verbindungen zu dem auf medizinische Inhalte spezialisierten Content- und Journal-Portal German Medical Science (GMS)<sup>16</sup> werden auch vor dem Hintergrund neuer Rahmenbedingungen von beiden Seiten intensiviert und mögliche Synergieeffekte, z. B. im Bereich der Ressourcenplanung und der Geschäftsmodelle, weiter erschlossen.

## 3 Fazit und Ausblick

Die DiPP-Initiative hat sich nach der ersten Phase ihrer Etablierung als Plattform für elektronische Open Access-Zeitschriften als Dienstleister positioniert und konnte im Hinblick auf kooperierende Institutionen, Nutzungszahlen und veröffentlichte Dokumente expandieren. Die Dokumententwicklung der abgedeckten Fachdisziplinen (vgl. Abb. 1) zeigt einen für den Open Access-Bereich eher singulären Zuwachs im Sektor der Geistes- und Sozialwissenschaften. Dieser erfreuliche Sachverhalt hat jedoch Konsequenzen für die Konzeption künftiger Geschäftsmodelle, die sich auch – wie das Beispiel von GMS beweist – am finanziellen Hintergrund der entsprechenden *scientific communities* bzw. Fachgesellschaften orientieren kann. Über konzeptionelle Überlegungen hinausgehend existieren die oben ge-

---

<sup>15</sup> URL: [www.elanguage.net](http://www.elanguage.net)

<sup>16</sup> URL: <http://www.egms.de/de/>



nannten konkreten Ansätze, Möglichkeiten zur Erwirtschaftung von Einnahmen zu erproben. Diese Maßnahmen weisen bereits darauf hin, dass die nächsten Schritte zur Sicherung der Nachhaltigkeit im weiteren Ausbau der Zusammenarbeit mit den lokalen Zeitschriftenredaktionen und mit externen Partnern zu sehen sind.

## 4 Literatur und Links

Alle Links wurden zuletzt am 4.4.2007 besucht.

Bauer, Bruno: „Kommerzielle Open Access Publishing-Geschäftsmodelle auf dem Prüfstand: ökonomische Zwischenbilanz der Gold Road to Open Access an drei österreichischen Universitäten.“ GMS Med Bibl Inf 2006;6(3):Doc32.

<http://www.egms.de/en/journals/mbi/2007-6/mbi000050.shtml>

Deutsche Initiative für Netzwerkinformation e.V., DINI-Zertifikat 2007:

[http://www.dini.de/documents/dini\\_zertifikat\\_2007.pdf](http://www.dini.de/documents/dini_zertifikat_2007.pdf)

elanguage.net:

[www.elanguage.net](http://www.elanguage.net)

eSciDoc:

<http://www.esdoc-project.de/de/materialien-downloads.html>

German Medical Science:

<http://www.egms.de/de/>

Horstmann, Wolfram: „Kooperationsmodelle für Open Access eJournals in der Publikationsinitiative DiPP NRW“, in : Knowledge eXtended (3. Konferenz der Zentralbibliothek), Schriften des Forschungszentrums Jülich, Reihe Bibliothek, Band 14 (2005), S. 109-122.

Horstmann, Wolfram und Jochen Schirrwagen, Peter Reimer: „Multi-level eJournal support structures in the initiative Digital Peer Publishing NRW.“ Joint Workshop on Electronic Publishing, Organised by Delos, SVEP and ScieCom. Lund University Libraries, Sweden, 2005.

[http://www.dipp.nrw.de/ueberdipp/publikationen/horstmann\\_lund\\_full.pdf](http://www.dipp.nrw.de/ueberdipp/publikationen/horstmann_lund_full.pdf)

Institut für Rechtsfragen der Freien und Open Source Software:

<http://www.ifross.de/>

International Journal of Conflict and Violence

<http://www.ijcv.org/>

Knowledge eXtended (3. Konferenz der Zentralbibliothek), Schriften des Forschungszentrums Jülich, Reihe Bibliothek, Band 14 (2005).

<http://juwel.fz-juelich.de:8080/dspace122fzj/handle/2128/396>

Kompetenznetzwerk für Bibliotheken: Bibliotheksportal:

<http://www.bibliotheksportal.de/hauptmenue/themen/digitale-bibliothek/elektronisches-publizieren/akteure-und-projekte>

Neumann, Jan: „Auf dem Weg zu einem Open-Access-Geschäftsmodell“. In: Open Source Jahrbuch 2006, Kapitel Wissenschaft<sup>2</sup>: Open Access.

[http://www.opensourcejahrbuch.de/download/jb2006/chapter\\_07/osjb2006-07-01-neumann](http://www.opensourcejahrbuch.de/download/jb2006/chapter_07/osjb2006-07-01-neumann)

Nötzelmann, Cordula und Sören Lorenz: „Das Redaktionstandem als innovatives Kooperationsmodell zwischen Fachwissenschaftlern und Bibliothekaren am Beispiel des Open Access E-Journals Brains, Minds & Media“, Knowledge eXtended (3. Konferenz der Zentralbibliothek), Schriften des Forschungszentrums Jülich, Reihe Bibliothek, Band 14 (2005), S. 91-100.

Universität Bielefeld, Übersichtsseite zur Neuausrichtung im Wissenschaftlichen Publizieren und Open Access:

<http://www.uni-bielefeld.de/ub/wp/projekt.htm>

University of Queensland, Institutional Repository:

<http://espace.library.uq.edu.au/>

Wissensmanagement. Zeitschrift für Innovation, 1/2006, Lemmens Verlag, Bonn :

[http://www.lemmens.de/verlag/zeitschriften/wima\\_special\\_aktuell/special1\\_06.pdf](http://www.lemmens.de/verlag/zeitschriften/wima_special_aktuell/special1_06.pdf)

# **Erfahrungen mit Community-Funktionen im PsychLinker\***

*Björn Mittelsdorf & Ute Wahner*

Zentrum für Psychologische Information und Dokumentation (ZPID)

## **Zusammenfassung**

Der Beitrag vermittelt Erfahrungen, die das ZPID mit dem PsychLinker seit 1996 sammeln konnte. Von Beginn an bestand die Konzeption des PsychLinker in weiten Teilen aus Funktionen, die heute unter Schlagworten wie Web 2.0, Computer Supported Collaborative Work und Online-Community besprochen werden.

Die Integration dieser neuen Techniken in den Wissenschaftsbetrieb wird als unvermeidlich betrachtet. Auf unseren Erfahrungen aufbauend, können neue Projekte in diesem Bereich bereits in der Konzeptionsphase einige Fallstricke vermeiden.

## **1 Beschreibung des PsychLinker**

Das ZPID betreut mit dem PsychLinker seit 1996 einen stetig wachsenden und qualitativ hochwertigen Linkkatalog.

Der Erfolg des PsychLinker lässt sich u. a. auf zwei weit verbreitete Methoden zurückführen:

- Zusammenfassung räumlich getrennter, thematisch zusammengehöriger Linksammlungen
- Zweistufiges Vorschlagsmodell für neue Webressourcen

Die flache Hierarchie des WWW begünstigt den Einzelnen dabei, interessante Quellen zu erzeugen und zu verbinden. Gleichzeitig sind viele dieser Angebote zumindest teilweise redundant.

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 479-483.

Im Projekt Digitale Psychologie Information (DPI)<sup>1</sup> wurden die beim ZPID vorliegenden Links mit denen des FIPS<sup>2</sup> und denen des Portals von *psychologie.de* nach einer Dubletten- und Qualitätskontrolle kombiniert. Dabei wurden – im Gegensatz zu vielen anderen Angeboten im Internet – die entsprechenden Daten nicht einfach plagiiert; stattdessen wird deutlich auf die Herkunft und das Engagement der Beteiligten hingewiesen. Dies erfolgt ebenso bei Linksammlungen, die dem ZPID von einzelnen Wissenschaftlerinnen und Wissenschaftlern zur weiteren Pflege übergeben werden. Das gute Suchmaschinen-Ranking, das der PsychLinker mittlerweile vorweisen kann, führt zu einer noch höheren Sichtbarkeit der geleisteten Arbeit. Hierfür wurde auf viele Gestaltungsmöglichkeiten bewusst verzichtet und zusätzlich mithilfe von Metatags die Interpretierbarkeit des Psychlinker durch Suchmaschinen-Robots optimiert.

Betrieb, Betreuung und technische Administration liegen in den Händen des ZPID. Ein Teil der Rubriken wird von Mitarbeitern von ZPID und SULB betreut. Erweiterungen des PsychLinker erfolgen auf folgenden Wegen:

- „Normale“ Besucher der Seite können – ohne sich zu registrieren – Vorschläge an die Redaktion senden.
- Personen mit fachlicher Expertise stellen dem ZPID eine von ihnen erstellte Linksammlung zur Verfügung, um die dauerhafte Pflege und Veröffentlichung zu gewährleisten.
- Fachredakteure haben sich zur kontinuierlichen Mitarbeit am PsychLinker in einem thematisch umgrenzten Gebiet bereiterklärt, auf dem sie fachliche Expertise besitzen. Ein Redakteur ist für mindestens eine Kategorie zuständig; dies erlaubt ihm, neue Verweise in dieser Kategorie freizuschalten. Diese wurden entweder von ihm selbst recherchiert, oder er wurde durch das System auf einen Vorschlag aufmerksam gemacht. Redakteure sind für die Metadaten URL, Titel (automatische Zuspiegelung) und Kurzbeschreibung des Inhalts zuständig. Die Vergabe weiterer Metadaten (angelehnt an DCMI<sup>3</sup>) ist fakultativ; fehlende Metadaten werden durch ZPID ergänzt.

---

<sup>1</sup> Drittmittelprojekt der Saarländischen Universitäts- und Landesbibliothek (SULB) und des Zentrums für Psychologische Information und Dokumentation (ZPID) mit Unterstützung der Deutschen Forschungsgemeinschaft (DFG). Leitung: Bernd Hagenau und Erich Weichselgartner.

<sup>2</sup> Fachinformationsführer Psychologie (<http://fips.sulb.uni-saarland.de/intern/index.htm>).

<sup>3</sup> Dublin Core Metadata Initiative (<http://dublincore.org>).

## **2 Community-Funktionen im PsychLinker**

Der PsychLinker kann für sich in Anspruch nehmen, schon seit langer Zeit Community-Funktionen anzubieten, so dass das ZPID bereits vielschichtige Erfahrungen im Bereich der neuen Netzgesellschaft sammeln konnte. Ähnlich wie bei Wikipedia kann jeder Internetbewohner Vorschläge für den Linkkatalog einreichen, diese werden dann einem Fachredakteur des PsychLinker angezeigt und entweder verworfen oder – nach Vergabe der Metadaten – freigeschaltet. Diese Mischung aus unautorisierten und registrierten Usern sowie manuellen und automatischen Prozessen vereint Vorteile und kompensiert Nachteile, u a. die folgenden:

- Schutz vor Board-Spamming
- niedrige Hemmschwelle für neue Vorschläge
- intellektuelle Prüfung neuer Vorschläge, etwa auf Aktualität und Richtigkeit des Inhalts, auf Redundanz etc.

Offene Systeme, wie zum Beispiel Wikipedia oder Internetforen, sind ständig automatisierten und individuellen Link-Spamming-Angriffen ausgesetzt. Der Grundgedanke des freien Zugangs beinhaltet leider auch den Zugang für Teilnehmer mit unlauteren Absichten. Diese benutzen fremde Foren, um Werbung für eigene Angebote zu machen, mit der Seriosität des ausgenutzten Forums das eigene Suchmaschinen-Ranking zu verbessern oder auf einen manipulierten Webspaces zu verweisen, um Phishing oder Cross-Site-Scripting zu betreiben. Automatische Filter, Blacklists, Catchpas u. s. w. können niemals eine vollständige Sicherheit garantieren. Sie stellen zudem teilweise eine erhebliche Hürde für reguläre Nutzer dar, die dann u. U. frustriert nicht mehr zum Gemeinschaftsprojekt beitragen.

Eine intellektuelle Prüfung durch thematisch eingedachte Personen ist die Voraussetzung dafür, die Qualität der nachgewiesenen Links und die der vergebenen Metadaten zu sichern. Durch eine Verteilung der Arbeit auf mehrere Schultern reduziert sich der Aufwand für die einzelnen Mitwirkenden.

Fachredakteure sind sowohl in der Lage, Aktualität und inhaltliche Richtigkeit eines Informationsangebotes zu beurteilen, sie können vielfach aufgrund ihrer fachlichen Expertise auch validere Metadaten vergeben, etwa Kurzbeschreibungen oder kontrollierte Schlagworte. Durch die Kooperation mit ausgewählten Fachexperten und den durch qualitativ hochwertige Metadaten erzielten Mehrwert ist der PsychLinker als fachliche Linksammlung allgemeinen Angeboten, wie zum Beispiel [www.dmoz.org](http://www.dmoz.org), überlegen.

Derzeit engagieren sich 14 externe, eine SULB-interne und sechs ZPID-interne Redakteure im PsychLinker. Seit 2002 sind vier Redakteure aus beruflichen Gründen ausgeschieden, im gleichen Zeitraum sind sechs neue Redakteure hinzugekommen. Mit einer Ausnahme sind die externen Fachredakteure Psychologinnen und Psychologen, die an Universitäten oder anderen Forschungsinstituten tätig sind.

Zwei Fachredakteure wurden aus dem Team von psychologie.de übernommen, die Mehrzahl meldete sich auf eine Rundmail an alle Mitglieder der Deutschen Gesellschaft für Psychologie im Jahr 2001. Das ZPID wendet sich regelmäßig bei Tagungen und Kongressen an die Fachöffentlichkeit mit der Bitte um Mitwirkung. Es spricht auch gezielt einzelne Personen an, die es als potentiell interessiert und geeignet betrachtet.

Ein Anschreiben an die Psychologen unter den Fachreferenten der Universitätsbibliotheken erbrachte keine positive Resonanz. Als Gründe für die Nichtbeteiligung wurden Zeitmangel und Unsicherheit hinsichtlich der inhaltlichen Expertise angegeben.

Als Redakteur verpflichtet man sich, mindestens einmal im Jahr die Beschreibungen der Links der eigenen Kategorien auf Aktualität zu prüfen. Bei Eingabe neuer Ressourcen reichen URL, Titel und Beschreibung, wobei das System auf Basis der Meta-Tags Vorschläge unterbreitet.

Der Zeitaufwand für die Fachredakteure ist u. a. abhängig vom Umfang der Linksammlung, den persönlichen Fähigkeiten zur Inhaltszusammenfassung, der Aktualisierungs- und Erweiterungshäufigkeit und von der Menge der vergebenen Kernmetadaten. Zudem macht es natürlich auch einen Unterschied, ob jemand seine bereits vorhandene Bookmarkliste einpflegt, oder aber ob eine neue Linksammlung erstellt wird.

Die Fachredakteure werden bei ihrer Arbeit auf verschiedene Weise vom ZPID unterstützt: Ihnen steht ein Handbuch zur Verfügung, das u. a. Hinweise zur Auswahl und Dokumentation von Links enthält und in die Arbeit mit dem Redaktionssystem einführt. Ferner werden sie zu Beginn ihrer Mitwirkung und später bei Bedarf per E-Mail oder telefonisch beraten. Diese Maßnahmen dienen gleichzeitig der Qualitätskontrolle des Angebotes.

Die zeitliche Belastung der Redakteure wird auch dadurch in Grenzen gehalten, dass fehlende Metadaten vom ZPID eingetragen und gepflegt werden. Zudem übernimmt das ZPID die in kurzen zeitlichen Abständen erfolgende Prüfung der Erreichbarkeit der Links.

Auch wenn die Fachredakteure als Dank vom ZPID einen kostenfreien Zugang zur Datenbank PSYNDEX erhalten, so überwiegen bei der Mitwirkung am PsychLinker doch die intrinsischen Motivationen: Die Fachredakteure sind sich bewusst, dass sie einen Beitrag zur adäquaten Darstellung der Psychologie in der Öffentlichkeit und zur Verbesserung der Informationsversorgung leisten, sie schätzen die Möglichkeit, an einem nützlichen Gemeinschaftsprojekt mitzuwirken und knüpfen Kontakte; sie profitieren davon, ihr Forschungsthema in der Öffentlichkeit darstellen und es mit dem eigenen Namen verknüpfen zu können, zusätzlich sind auch Einführungstexte möglich. Die Fachredakteure berichten zudem von eigenem Informationsgewinn und von ihrer Freude am Recherchieren und Aufbereiten von Fachinformationen.

Das Gelingen eines solchen Gemeinschaftsprojektes hängt nach unseren Erfahrungen entscheidend von folgenden Faktoren ab:

- Gute Anbindung an die Fachcommunity, um bei der Rekrutierung hoch qualifizierter Fachredakteure erfolgreich zu sein
- komfortables und sicheres Redaktionssystem; Betreuung der Fachredakteure
- begrenzter Metadatensatz
- konstante Mitarbeit von bezahltem, fachlich und dokumentarisch geschultem Personal

### **3 Ausblick**

Der PsychLinker, dessen Design-Schwerpunkte auf Schlichtheit und Barrierefreiheit liegen, wird derzeit mit neuen Funktionen angereichert. Push-Dienste wie zum Beispiel RSS-Feeds und eine noch effizientere Linkprüfung werden auf ihren Nutzen untersucht.

Die Anwerbung weiterer Fachredakteure wird nach Abschluss der technischen Arbeiten verstärkt erfolgen.





# Eine Informationsarchitektur für wissenschaftliche Fachportale in vascoda\*

*Sabine Heinz & Maximilian Stempfhuber*

GESIS / Informationszentrum Sozialwissenschaften

Lennéstr. 30

53113 Bonn

*{heinz, stempfhuber}@iz-soz.de*

## **Zusammenfassung**

Die rund 40 im Wissenschaftsportal vascoda eingebundenen Fachportale bieten auf ihre jeweiligen Fächer und Fachcommunities abgestimmte Informationsangebote. Aus der interdisziplinären Sicht auf die Fachportale, die durch die Bündelung der Portale in vascoda erzeugt wird, fällt eine erhebliche strukturelle und semantische Heterogenität auf. Diese drückt sich u. a. in unterschiedlichen Informationsangeboten, Benennungen und Portalstrukturen aus und führt bei Nutzerinnen und Nutzern, die in mehreren Fachportalen recherchieren, zu einer hohen kognitiven Last und erschwerten Bedienbarkeit. Durch die Entwicklung eines einheitlichen Modells der Informationsarchitektur für die wissenschaftlichen Fachportale soll die Grundlage für eine strukturelle Angleichung und damit eine einheitliche Gestaltung auf ergonomischer und ästhetischer Ebene geschaffen werden.

## **I Einleitung**

Im Wissenschaftsportal vascoda<sup>1</sup>, das vom Bundesministerium für Bildung und Forschung (BMBF) und der Deutschen Forschungsgemeinschaft (DFG) in mehreren (Teil-)Projekten<sup>2</sup> gefördert wird, sind derzeit rund 40 Fachportale aus den un-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 485-507.

<sup>1</sup> <http://www.vascoda.de>

<sup>2</sup> Der vorliegende Beitrag entstand im Rahmen des Projektes „Einheitliche Gestaltung der Fachportale und des vascoda-Portals“, Teilprojekt des „vascoda Themenblocks 5: Fachportale – Übergreifende Strukturen und Aufgaben“, der von der DFG unter dem Geschäftszeichen 554 9022(2) ZBW Kiel gefördert wird.

terschiedlichsten Wissenschaftsdisziplinen eingebunden. Verantwortet werden die Fachportale von derzeit 39 Mitgliedsinstitutionen<sup>3</sup> des vascoda-Vereins.

Bei einer vergleichenden Betrachtung der einzelnen Informationsangebote, die von den Fachportalen bereitgestellt werden, fällt ein hohes Maß an struktureller und semantischer Heterogenität ins Auge. Zurückzuführen ist dies darauf, dass die Angebote in den Fachportalen auf das jeweilige Fach und die Fachcommunity abgestimmt sind. In einer im Auftrag der DFG durchgeführten Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung wurde festgestellt, dass sich die NutzerInnen einerseits eine domänenspezifische und tiefe Erschließung in den Fachportalen wünschen, andererseits aber auch Verbindungen zwischen den Fächern/Domänen (Clustern), die die interdisziplinäre Forschung unterstützen. Als besonderes Problem interdisziplinärer Forschung wird der erhöhte Zeitaufwand von Informationsrecherche und –beschaffung identifiziert (vgl. Deutsche Forschungsgemeinschaft 2003). Zurzeit müssen interdisziplinär in mehreren Fachportalen recherchierende NutzerInnen eine hohe kognitive Last und erschwerte Bedienbarkeit beim Wechsel zwischen den Angeboten in Kauf nehmen: Dadurch, dass gleiche Informationsangebote in den Fachportalen unterschiedlich benannt und/oder an unterschiedlichen Stellen im Portal angeordnet sind, erhöht sich zwangsläufig der Lernaufwand für den Nutzer, der Transfer von Altwissen wird erschwert und die Bildung eines für alle Fachportale einheitlichen mentalen Modells behindert. Ein solches wird jedoch benötigt, um Beziehungen zwischen den Informationseinheiten identifizieren zu können (vgl. Germonprez & Zigurs 2003, 10). Auch Wirth (2004) weist auf die Bedeutung eines mentalen Modells zur Navigation in komplexen Websites hin.

Im November 2006 wurde im Auftrag des vascoda-Qualitätsmanagements von der Göttinger Firma eResult ein Usability-Test des vascoda-Portals mit 24 Testpersonen durchgeführt. Ein Teil des Tests zielte darauf ab, die Testpersonen den Übergang von vascoda in die Fachportale, also von der Trefferliste in vascoda zur Einzeltrefferanzeige im Design des jeweiligen Fachportals, bewerten zu lassen. Die Mehrheit der Testpersonen gab an, dass für sie ein einheitlicher Aufbau der Trefferanzeigen wichtig bzw. sehr wichtig ist (vgl. ZB MED 2006, 161-172). Dies kann als Hinweis darauf gewertet werden, dass ein einheitlicher Aufbau der Fachportale zumindest für zentrale, häufig genutzte Angebote und Funktionalitäten

---

<sup>3</sup> Eine Liste der Mitglieder findet sich hier:  
[http://www.vascoda.de/Mitglieder\\_vascoda\\_Logo-de.pdf](http://www.vascoda.de/Mitglieder_vascoda_Logo-de.pdf)

wünschenswert ist, um die Bildung eines gemeinsamen mentalen Modells und somit ein schnelles Zurechtfinden des Nutzers zu ermöglichen.

Im Rahmen des Projekts „Einheitliche Gestaltung der Fachportale und des vascoda-Portals“, Teilprojekt des „vascoda Themenblocks 5: Fachportale – Übergreifende Strukturen und Aufgaben“, soll die Grundlage für eine mit vertretbarem Aufwand durchzuführende strukturelle Angleichung und einheitliche Gestaltung der Fachportale auf ergonomischer und ästhetischer Ebene geschaffen werden. Dazu bedarf es zunächst einer einheitlichen Informationsarchitektur der Portale, die in einem weiteren Schritt in eine konsistente Oberflächengestaltung umgesetzt werden soll. Diese Oberflächengestaltung soll auf einer gemeinsamen Formensprache basieren und gleichzeitig den Fachportalen genügend eigenen Gestaltungsspielraum bieten. Dabei soll ein komplettes Gestaltungs- und Bedienkonzept unter Berücksichtigung softwareergonomischer Richtlinien entstehen. Das Ziel ist, die Fachportale so an die Bedürfnisse des Nutzers/der Nutzerin anzupassen, dass ein einheitliches Look & Feel entsteht und ein größtmögliches Maß an Gebrauchstauglichkeit (engl. *Usability*) erreicht wird. Dabei bezieht sich Gebrauchstauglichkeit auf die drei Kriterien Effektivität, Effizienz und Nutzerzufriedenheit.<sup>4</sup>

Im Folgenden werden zunächst der Begriff der Informationsarchitektur, das dahinter stehende Konzept und seine Anwendung auf den Kontext der Fachportale näher erläutert. Anschließend wird der Status Quo in den Fachportalen beschrieben. Danach wird ein Referenzmodell für eine Informationsarchitektur der Fachportale entwickelt. Das abschließende Kapitel gibt einen Ausblick darauf, wie das Modell exemplarisch in einem Musterportal umgesetzt werden soll.

## **2 Informationsarchitektur**

### **2.1 Begriffsklärung und Definition**

Der Begriff Informationsarchitektur oder englisch *Information Architecture* (IA) ist in erster Linie im US-amerikanischen Raum seit Ende der 1990er Jahre bekannt und gebräuchlich. Wegweisend bei der Einführung des Begriffs war das 1998 erschienene Buch *Information architecture for the World Wide Web: designing large-scale Web sites* der beiden Bibliotheks- und Informationswissenschaftler Rosenfeld und Morville. Vor dem Hintergrund des rasant wachsenden World Wide Web

---

<sup>4</sup> Zu den Grundlagen der Softwareergonomie vgl. Balzert et al. (1988), Herczeg (1994) und Shneiderman (1998). Einen Überblicksartikel liefert Krause (2004).

definierten die Autoren zum ersten Mal das Konzept der Informationsarchitektur im Kontext von Websites. Sie übernahmen dabei den Begriff *information architecture* von dem Architekten und Grafikdesigner Richard Saul Wurman, der ihn bereits 1976 auf einer Konferenz des *American Institute of Architecture* benutzt hatte. 1996 veröffentlichte Wurman zusammen mit Peter Bradford das Buch *Information architects*, in dem er Informationsarchitektur in erster Linie als Präsentation und Layout von Informationen behandelt (vgl. Arndt 2006, 13ff.).

Morville und Rosenfeld grenzen sich von diesem Informationsarchitektur-Begriff ab; was Wurman als Informationsarchitektur auffasst, bezeichnen sie als *Information design* und definieren ihrerseits *Informationsarchitektur* wie folgt:

„information architecture n.

The structural design of shared information environments.

The combination of organization, labeling, search, and navigation systems within web sites and intranets.

The art and science of shaping information products and experiences to support usability and findability.

An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape.” (Morville & Rosenfeld 2006, 4)<sup>5</sup>

Informationsarchitektur beschäftigt sich also laut Morville und Rosenfeld mit der strukturellen Gestaltung von Informationssystemen. Im Einzelnen beinhaltet dies die Organisation, Benennung, Suche und Navigation innerhalb von Websites und Intranets – mit dem übergeordneten Ziel, die Gebrauchstauglichkeit dieser Informationssysteme zu erhöhen und Informationen leicht auffindbar zu machen. Dabei sollen Prinzipien aus dem Bereich des Designs und der Architektur angewandt werden (vgl. Morville & Rosenfeld 2006, 10f.).

Der Ansatz, den Morville und Rosenfeld verfolgen, ist in erster Linie praxisorientiert: Sie verstehen ihr Buch als schrittweise Anleitung zur Konzeptionierung und Erstellung von Websites, indem sie z. B. in Methoden wie *Content Inventory*, *Card Sorting*, *Wireframes*, *Prototyping* etc. einführen. Gleichzeitig trugen sie mit der Veröffentlichung ihres Buches dazu bei, das Berufsbild und die Arbeitsfelder eines Informationsarchitekten (*information architect*) zu definieren und zu etablieren – zumindest für den US-amerikanischen Raum. Im Jahr 2002 gründeten sie schließlich zusammen mit Christina Wodtke und anderen das *Information Architecture*

---

<sup>5</sup> Diese Definition findet sich in der dritten, völlig überarbeiteten und erweiterten Auflage ihres Buches von 1998.

*Institute*<sup>6</sup>, das es sich u a. zum Ziel gesetzt hat, Informationsarchitektur als Berufsfeld bekannt zu machen und zu fördern.

Seit 2000 findet einmal jährlich in den USA der *Information Architecture Summit*<sup>7</sup> statt, organisiert von der *American Society for Information Science and Technology* (ASIS&T). Das Thema des IA Summit 2006 lautete „Learning, Doing, Selling“, 2007 wird sich die Konferenz mit „Enriching IA – rich information, rich interaction, rich relationships“ beschäftigen. 2005 fand zum ersten Mal ein europäischer IA Summit statt, gefolgt von einer zweiten Konferenz 2006<sup>8</sup>, ebenfalls initiiert und unterstützt von ASIS&T. Die Themenblöcke und Beiträge auf den IA Summits haben vorrangig Praxisbezug.

Der geschilderte stark praxisorientierte Ansatz wird in der IA-Community auch kritisch gesehen: So beklagt Fast (2006) einen Mangel an wissenschaftlicher Grundlagenforschung im Bereich der Informationsarchitektur. Dies sei in der Geschichte der Disziplin begründet, da Informationsarchitektur von Praktikern konzipiert, definiert und entwickelt worden sei. Er weist jedoch auch darauf hin, dass es Forschung im Bereich Informationsarchitektur durchaus gibt, so z. B. nachzulesen im *Journal of the American Society for Information Science and Technology* (JASIST), Heft 10 (53), einem *special issue* zur Informationsarchitektur. Die Heftbeiträge befassen sich sowohl mit theoretischen Konzepten in der Informationsarchitektur als auch mit Fallstudien aus der Praxis (vgl. JASIST 2002).

Morville und Rosenfeld sehen Überschneidungen der Informationsarchitektur u a. zu folgenden wissenschaftlichen Disziplinen: Grafikdesign, Interaction Design, Usability Engineering, Softwareentwicklung und Wissensmanagement (vgl. Morville & Rosenfeld 2006, 10f.). Gerade diese Überschneidungen spielen bei der Entwicklung des Referenzmodells für die in vascoda eingebundenen Fachportale eine Rolle.

## **2.2 Informationsarchitektur im Kontext der Fachportale**

Aufgrund der Überschneidungen zu den genannten Fachgebieten eignet sich der von Morville und Rosenfeld definierte Begriff der Informationsarchitektur zur Übertragung auf den Kontext der Fachportale. Im Folgenden verstehen wir unter Informationsarchitektur im Sinne von Morville und Rosenfeld die strukturelle Gestaltung der Fachportale, die sich aus folgenden Komponenten zusammensetzt: (1)

---

<sup>6</sup> <http://iainstitute.org>

<sup>7</sup> <http://www.iasummit.org>

<sup>8</sup> <http://www.euroia.org/2005> bzw. <http://www.euroia.org>

Organisation/Strukturierung der Information, (2) Benennung (*Labeling*), (3) Suche und (4) Navigation. Erweitern wollen wir die Definition um die Ausrichtung auf die softwareergonomische Gestaltung, d.h. um die Frage, wie sich die strukturelle Gestaltung mit softwareergonomischen Mitteln in eine konsistente Oberfläche umsetzen lässt.

Unser Fokus bei der Entwicklung des Referenzmodells zur gemeinsamen Gestaltung der Fachportale in vascoda liegt im ersten Schritt auf der Organisation und Strukturierung der Information. Im Kern geht es darum, wie die Informationsangebote in den Fachportalen kategorisiert werden können. Dabei gilt es, folgende Fragen zu beantworten: Welche Angebote gibt es? Wodurch zeichnen sie sich aus bzw. wodurch grenzen sie sich voneinander ab? Welche Nutzerperspektiven auf die Angebote gibt es und wie können sie abgebildet werden? Welche Angebote können in welchen Kategorien gruppiert werden? Durch welche Kriterien zeichnen sich diese Kategorien aus? Welche Verknüpfungen gibt es zwischen einzelnen Kategorien/Angeboten? Wie soll mit der Heterogenität vermeintlich gleicher Angebote umgegangen werden?

Zunächst befassen wir uns also mit diesen grundlegenden Überlegungen zur Organisation und Strukturierung der in den Fachportalen angebotenen Information. Daran knüpfen Fragen der Benennung, der Suche und der Navigation innerhalb der Informationsarchitektur an, die an dieser Stelle jedoch nur angerissen werden sollen. Bei der Benennung der Inhalte geht es darum, ein kohärentes Bezeichnungssystem zu entwickeln, das sowohl die Inhalte adäquat repräsentiert als auch die Bedürfnisse der Nutzer berücksichtigt. Bei der Konzeption der Suche muss entschieden werden, welche Informationsangebote wie durchsuchbar sein sollen. Dazu gehören u. a. folgende Fragen: Welche Retrieval-Verfahren werden eingesetzt? Wie sieht die einfache Suche aus, wie die erweiterte? Welche Sucheinschränkungen werden ermöglicht? Wie sieht die Ergebnisliste aus? etc. Die Navigation im Fachportal baut stark auf der Organisation der Inhalte auf. Eine Menüstruktur zum Blättern im Portal kann sich unmittelbar aus der Inhaltsstruktur entwickeln. Bei der Konzeption der Navigation ist u. a. zu klären, wo die Hauptnavigation platziert wird, wie viele Menüebenen es gibt und welche Navigationselemente (z. B. Karteireiter, Breadcrumbs etc.) eingesetzt werden. Die Struktur der Inhalte bietet dabei einen Orientierungsrahmen und Wegweiser für die Navigation. Dabei gilt es, die notwendigen softwareergonomischen Mittel zu finden, um den NutzerInnen einen effizienten Zugriff auf die Inhalte zu ermöglichen. Mit den genannten Punkten wurden die komplexen Themen Suche und Navigation nur angerissen, da der Fokus des Referenzmodells der Fachportale auf der Organisation und Strukturierung der Inhalte

liegt und die ergonomischen/ästhetischen Aspekte vom Musterdesign, das parallel dazu entwickelt wird, abgedeckt werden.

### **3 Ist-Zustand der in vascoda eingebundenen Fachportale**

#### **3.1 Heterogene Angebotsstruktur**

Wie in mehreren Untersuchungen bereits festgestellt wurde (vgl. Mayr 2006, Pianos & Trauzold 2006, Heinz & Pianos 2007), zeichnen sich die Informationsangebote der Fachportale durch eine starke Heterogenität in struktureller sowie semantischer Hinsicht aus. Die Palette der Informationsangebote ist insgesamt sehr breit: Pianos & Trauzold haben in einer Befragung der Fachportal-Betreiber an die 20 unterschiedliche Angebote, im vascoda-Kontext Module genannt, identifiziert. Dabei handelt es sich u. a. um Fachinformationsführer (Nachweis von Internetquellen), Bibliothekskataloge (OPACs), Zeitschriftenaufsatz-Datenbanken, Fachdatenbanken und Volltextserver. Diese genannten Module werden jeweils von über 80% der Fachportale angeboten oder befinden sich in Planung. Andere Module wiederum sind nur in wenigen Portalen vorhanden bzw. geplant, z. B. Jobangebote (16%) oder Digitale Handbücher/Lexika (7%). Die folgende Tabelle gibt einen Überblick über das Vorkommen (ausgewählter) Module in 43 befragten Fachportalen:

<i>Modul</i>	<i>Wie viele Fachportale (in Prozent) bieten das Modul an bzw. planen es?</i>
Fachinformationsführer	96 %
Bibliothekskataloge	91 %
Fachdatenbanken	91 %
Zeitschriftenaufsatz-Datenbanken	83 %
Volltextserver	81 %
Neuerwerbungslisten	76 %
Zeitschriften-Datenbanken	57 %
Online-Tutorials	48 %
Expertendatenbanken	48 %
Veranstaltungskalender	35 %
Jobangebote	16 %
Online-Pressearchive	15 %
Digitale Handbücher/Lexika	7 %
Bibliometrie/Zitationsanalyse	4 %

*Tabelle 1: Module in den Fachportalen*

Folglich unterscheidet sich die Angebotsstruktur der Fachportale zum Teil erheblich: Während beispielsweise das geschichtswissenschaftliche Fachportal Clio-online<sup>9</sup> insgesamt 18 Module anbietet, sind es bei es bei der Virtuellen Fachbibliothek Ost- und Südostasien CrossAsia<sup>10</sup> sechs (wobei es sich nicht um fixe Zahlen handelt und die Anzahl der Module nichts über die Qualität der Angebote aussagt). Einzelne Module finden sich zwar in vielen Fachportalen wieder, unterscheiden sich aber häufig in mindestens einem der folgenden Punkte: Struktur, Inhalt, Benennung oder Position innerhalb des Fachportals. Dies lässt sich am Beispiel des Fachinformationsführers in den beiden Fachportalen Vlib-AAC: History<sup>11</sup> (Virtual Library of Anglo-American Culture: History) und baltica-net<sup>12</sup> (Virtuelle Fachbibliothek Kulturkreis Baltische Länder) zeigen: Während Vlib-AAC: History angibt, im Fachinformationsführer nur institutionelle und thematische Websites zu erfassen, werden bei baltica-net darüber hinaus noch elektronische Volltexte aufgenommen. Bei Vlib-AAC: History ist der Fachinformationsführer als „History Guide“ betitelt, bei baltica-net heißt er „Internetquellen“. In Vlib-AAC: History findet man den Fachinformationsführer als externen Link innerhalb des Menüpunkts „Internet“, in baltica-net als Untermenüpunkt von „Recherche“ im Portal integriert.

### 3.2 Heterogene Sichten auf die Informationsangebote

Die Art und Weise, wie die Fachportale ihre Angebote strukturieren und gruppieren, also welche Sicht sie den Nutzerinnen und Nutzern auf ihr Portal bieten, variiert – je nach dem, welches Informationsbedürfnis des Nutzers als das vorrangigste angesehen wird. Dabei lassen sich drei verschiedene Sichten identifizieren (die jedoch auch in Mischformen auftreten): Am häufigsten sind die Informationsangebote nach Informationstyp gruppiert (Beispiel ViFaOst – Virtuelle Fachbibliothek Osteuropa<sup>13</sup>: Einteilung in Bibliothekskataloge, Neuerwerbungen, Zeitschriften und Aufsätze, Internetressourcen, Fachdatenbanken und Bibliografien, Texte und Materialien, Fachservice). Darüber hinaus gibt es eine funktionale Sicht (Beispiel ViFaPol – Virtuelle Fachbibliothek Politikwissenschaft<sup>14</sup>: Einteilung in Systematische Suche und Übergreifende Suche) und eine fachliche (Beispiel Slavistik-

---

<sup>9</sup> <http://www.clio-online.de>

<sup>10</sup> <http://crossia.org>

<sup>11</sup> <http://www.sub.uni-goettingen.de/vlib/history/index.php>

<sup>12</sup> <http://www.baltica-net.de>

<sup>13</sup> <http://www.vifaost.de>

<sup>14</sup> <http://www.vifapol.de>



Portal<sup>15</sup>: Einteilung in Geschichte, Politik und Gesellschaft, Sprache und Literatur, Musik, Kunst, Fachübergreifend). Eine erwähnenswerte Kombination zweier Sichten findet man bei Propylaeum<sup>16</sup>, der Virtuellen Fachbibliothek für Altertumswissenschaften: Hier wird eine fachliche Sicht (Alle Fächer, Ägyptologie, Alte Geschichte, Klassische Archäologie, Klassische Philologie, Vor- und Frühgeschichte) mit einer nach Informationstypen geordneten (Bibliothekskataloge, Neuerwerbungen, Zeitschriften/Aufsätze, Fachdatenbanken, Digitale Sammlungen, Internetressourcen) verknüpft. Die folgende Tabelle fasst zusammen, wie viele von 40 untersuchten Fachportalen welche Sicht auf ihre Informationsangebote anbieten:

<i>Sicht auf das Informationsangebot</i>	<i>Anzahl der Fachportale</i>
nach Informationstyp	27
funktional	11
fachlich	1
fachlich UND nach Informationstyp	1

*Tabelle 2: Unterschiedliche Sichten auf das Informationsangebot in den Fachportalen*

Die verschiedenen Sichten und Inkonsistenzen innerhalb der Sichten drücken sich wiederum in der unterschiedlichen Gestaltung und Benennung der Menüs in den Fachportalen aus.

### 3.3 Heterogene Oberflächengestaltung

Neben der dargestellten Heterogenität auf der Ebene der Informationsangebote lassen sich noch auffällige Unterschiede bezüglich der Oberflächengestaltung der einzelnen Fachportale feststellen (vgl. Heinz & Pianos 2007). Das grafische Gesamtkonzept, Bildsprache und Designelementen der Fachportale unterscheiden sich ebenso wie Bildschirmaufteilung, Navigation, Menü- und Linkstruktur. Am Beispiel der Positionierung der Hauptnavigation illustrieren die folgenden vier Screenshots die Heterogenität im Design der Fachportale:

---

<sup>15</sup> <http://www.slavistik-portal.de> (vor dem Relaunch im März/April 2007)

<sup>16</sup> <http://www.propylaeum.de>



Abbildung 1: Hauptnavigation vertikal links – Beispiel Savifa<sup>17</sup>

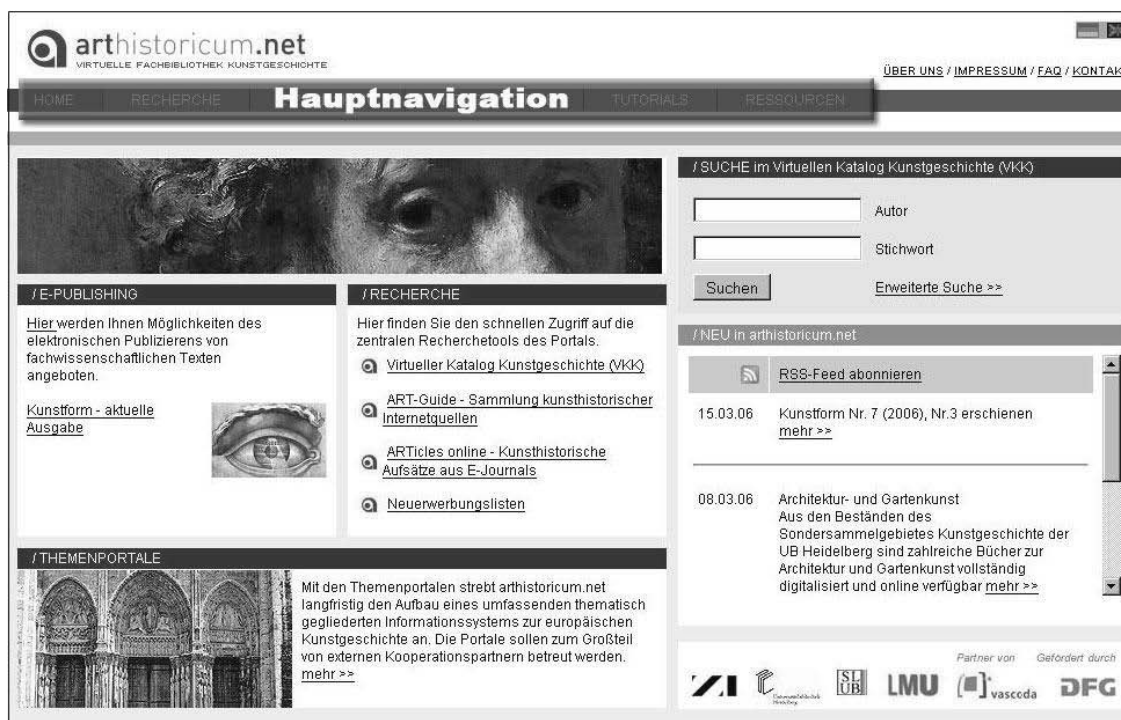


Abbildung 2: Hauptnavigation horizontal oben – Beispiel arthistoricum.net<sup>18</sup>

<sup>17</sup> Virtuelle Fachbibliothek Südasiens: <http://www.savifa.de>



Abbildung 3: Hauptnavigation vertikal links und horizontal oben – Germanistik imNetz<sup>19</sup>

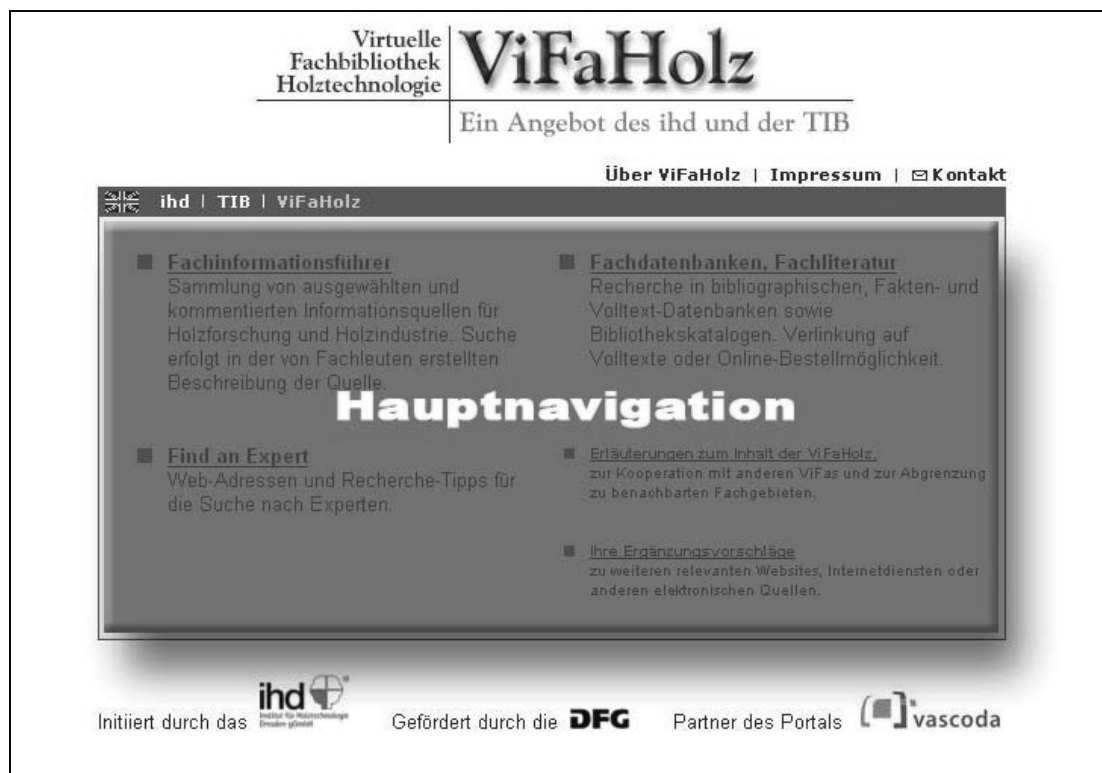


Abbildung 4: Hauptnavigation im Hauptbereich auf der Startseite – Beispiel ViFaHolz<sup>20</sup>

Die folgende Tabelle gibt einen Überblick darüber, wie viele Fachportale welche Form der Positionierung der Hauptnavigation gewählt haben (Stand: Sept. 2006):

<sup>18</sup> Virtuelle Fachbibliothek Kunstgeschichte: <http://www.arthistoricum.net>

<sup>19</sup> Virtuelle Fachbibliothek Germanistik: <http://www.germanistik-im-netz.de>

<sup>20</sup> Virtuelle Fachbibliothek Holz: <http://www.vifaholz.de>




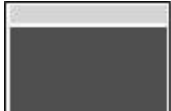
Positionierung der Hauptnavigation		Anzahl der Fachportale
	vertikal links	21
	horizontal oben	8
	vertikal links + horizontal oben	1
	im Hauptbereich auf der Startseite	7

Tabelle 3: Positionierung der Hauptnavigation in den Fachportalen

Die genannten Unterschiede zwischen den Fachportalen verdeutlichen, dass insgesamt keine einheitliche strukturelle Gestaltung, d.h. keine einheitliche Informationsarchitektur im oben definierten Sinne (vgl. Kap. 2.2) erkennbar ist. Es gibt vielmehr eine Spannbreite unterschiedlicher Modelle, die sich neben den Sichten auf die Informationsangebote vor allem in zwei Dimensionen unterscheiden: hinsichtlich der Inhalte und hinsichtlich des Integrationsgrads der Inhalte im Fachportal. Was die Inhalte anbetrifft, gibt es auf der einen Seite Fachportale, die ihren Schwerpunkt auf eigene redaktionelle Angebote (z. B. Themen, Tutorials etc.) legen (Beispiel: [historicum.net](http://www.historicum.net)<sup>21</sup>), und auf der anderen Seite Portale, die schwerpunktmäßig die verteilte Suche in Datenbanken organisieren (Beispiel: [infoconnex](http://www.infoconnex.de)<sup>22</sup>). Bezüglich des Integrationsgrads existieren einerseits Fachportale, die in erster Linie zu externen Informationsangeboten verlinken, und andererseits Fachportale, die verschiedene Informationsangebote unter einer Oberfläche im Portal integrieren (Beispiel: [ViFaPol](http://www.vifapol.de)<sup>23</sup>).

Begründet ist die dargestellte Spannbreite u. a. in der unterschiedlichen Entwicklungsgeschichte der Portale, den unterschiedlichen fachlichen Inhalten, den unterschiedlichen Zielgruppen sowie der unterschiedlichen Einbindung in institutionelle und nationale, teilweise auch internationale Zusammenhänge. Und auch wenn eine Unterscheidbarkeit von anderen Fachportalen sicherlich teilweise beabsichtigt ist,

<sup>21</sup> Geschichtswissenschaften im Internet: <http://www.historicum.net>

<sup>22</sup> Interdisziplinärer Informationsdienst für Pädagogik, Sozialwissenschaften und Psychologie : <http://www.infoconnex.de>

<sup>23</sup> Virtuelle Fachbibliothek Politikwissenschaft: <http://www.vifapol.de>

so stellt das Erscheinungsbild der Fachportale in seiner derzeitigen heterogenen Ausprägung eine vermeidbare Hürde beim Wechsel zwischen den Portalen dar. Ein wichtiger Schritt auf dem Weg zu mehr Nutzerfreundlichkeit ist deshalb eine gemeinsame Informationsarchitektur der Fachportale.

## **4 Ein Referenzmodell für die Fachportale in vascoda**

Um zukünftig eine einheitliche nutzerorientierte Sicht auf die Fachportale zu ermöglichen, stellen wir im Folgenden ein Referenzmodell vor, das die Grundlage für eine strukturelle Angleichung und übergreifende nutzerfreundlichere Gestaltung bietet. Das Referenzmodell umfasst eine für wissenschaftliche Portale optimierte Informationsarchitektur sowie ihre softwareergonomische Umsetzung in eine auf einem konsistenten Gestaltungs- und Bedienkonzept basierende Oberfläche.

### **4.1 Definition Referenzmodell**

Ein Referenzmodell erfüllt die Funktion eines Entwurfsmusters für eine Klasse von zu modellierenden Sachverhalten. Es bildet einen abstrakten Bezugsrahmen zur Definition von Entitäten und ihrer Beziehungen zueinander und ist dabei unabhängig von spezifischen Standards, Technologien, Implementierungen etc. Auf der Basis eines abstrakten Referenzmodells können spezifische Umsetzungen dieses Modells entwickelt werden (vgl. MacKenzie, M. et al. 2006). Bekannte Referenzmodelle im Bereich der Informatik sind das OSI-Schichtenmodell und das TCP/IP-Referenzmodell zur Beschreibung von Netzwerken.

Im Kontext digitaler Bibliotheken existieren unterschiedliche Ansätze für Referenzmodelle. Im *Digital Library Manifesto* des *DELOS Network of Excellence on Digital Libraries* wird explizit auf die Notwendigkeit eines Referenzmodells für digitale Bibliotheken hingewiesen:

Digital libraries need to obtain a corresponding Reference Model in order to consolidate the diversity of existing approaches into a cohesive and consistent whole, to offer a mechanism for enabling the comparison of different DLs, to provide a common basis for communication within the DL community, and to help focus further advancement“ (Candela et al. 2007).

Im *Manifesto* wird ein komplexes *Digital Library Universe* definiert, das aus drei Schichten (*Digital Library*, *Digital Library System*, *Digital Library Management System*) besteht, auf sechs Kernkonzepten (*Content*, *User*, *Functionality*, *Quality*, *Policy and Architecture*) beruht und in dem vier Akteursrollen (*DL End-Users*, *DL Designers*, *DL System Administrators*, *DL Application Developers*) identifiziert wer-

den. Das Modell orientiert sich also an den Mitteln der Softwarearchitektur, indem es Schichten, Komponenten und Beziehungen zwischen Komponenten definiert. Eine andere Form des Referenzmodells für digitale Bibliotheken stellt das 5S-Modell dar (vgl. Gonçalves et al. 2004). 5S steht für *Streams, Structures, Spaces, Scenarios, Societies*. Hierbei handelt es sich um ein formales Modell, das eine formale Sprache zur Beschreibung und Generierung digitaler Bibliotheken beinhaltet und als Ziel Vollständigkeit und Widerspruchsfreiheit verfolgt.

In beiden Modellen spielt die softwareergonomische Ebene keine explizite Rolle. Unser Ziel ist es deshalb, ein Modell zu entwickeln, das die konkrete inhaltliche Ebene der Fachportale berücksichtigt und mit der softwareergonomischen Ebene verknüpft. Während es bei der inhaltlichen Ebene um die Organisation und Strukturierung der Informationsangebote geht, fokussiert die softwareergonomische Ebene auf eine an größtmöglicher Gebrauchstauglichkeit orientierte Oberflächen-gestaltung, die einen effizienten Zugriff auf die Informationsangebote gewährleistet. Als dritte Ebene ist die technische Ebene zu erwähnen, die die zur Integration der Inhalte erforderliche technische Infrastruktur umfasst.<sup>24</sup> Im Folgenden wird zunächst die inhaltliche und anschließend die softwareergonomische Ebene beschrieben.

#### 4.2 Eine Informationsarchitektur für die Fachportale in vascoda

Grundlage für die Entwicklung des Referenzmodells auf inhaltlicher Ebene war eine Bestandsaufnahme zur strukturellen Gestaltung der Fachportale, die sich an folgenden Fragen orientierte: Welche Informationsangebote gibt es in welchen Fachportalen? Wie sind sie benannt? An welcher Stelle innerhalb der Gesamtstruktur des jeweiligen Portals sind sie angesiedelt? Dabei wurden auch die Umfrageergebnisse von Pianos & Trauzold (2006) einbezogen, darüber hinaus die Ergebnisse von Mayr (2006) sowie die im Rahmen des vascoda-Projekts „Metadaten, Standards, Heterogenitätsbehandlung“ in Zusammenarbeit mit der gleichnamigen vascoda-Arbeitsgruppe entwickelten vascoda-Sammlungsbeschreibungen<sup>25</sup> (engl. *Collection Level Descriptions*) und das von der vascoda-Facharbeitsgruppe Geisteswis-

---

<sup>24</sup> Eine wichtige Komponente der technischen Infrastruktur wird ein (Web-)Content-Management-System (CMS) zur Verwaltung der gesamten Oberflächenstruktur des Portals sein. Mit einem solchen System werden die Verwaltung und Pflege des Layouts und des Inhalts der statischen Seiten erleichtert. Ggf. kann über das CMS auch die Rechteverwaltung realisiert werden. In das CMS müssen dann die dynamischen Inhaltskomponenten integriert werden, also die Suche, das Browsing etc.

<sup>25</sup> <http://cld.vascoda.de/info/cluster>

senschaften definierte „Core-Set für eine geisteswissenschaftlich und/oder regional orientierte Virtuelle Fachbibliothek“<sup>26</sup>. Bei der zusammenführenden Betrachtung all dieser Ergebnisse fällt die komplexe und unübersichtliche Struktur der Informationsangebote auf, für die vor allem die unterschiedlichen Sichten auf die Angebote (vgl. Kap. 3.2) verantwortlich sind. Das Ziel war nun, die beschreibende Sicht zu verlassen und eine Modell- und Gruppenbildung der Informationsangebote in den Fachportalen vorzunehmen, und zwar anhand von intuitiven Kriterien, die in erster Linie die Nutzerperspektive berücksichtigen. Die Kriterien lassen sich wie folgt umreißen:

- Primäre Interaktion  
Wie erfolgt die primäre Interaktion des Nutzers/der Nutzerin mit dem Informationsangebot?  
z. B. Suchen, Blättern, Anfragen, Feedback/ Austausch <sup>27</sup>
- Informationsbedürfnis  
Welches Informationsbedürfnis des Nutzers/der Nutzerin liegt vor?  
z. B. thematisch (gesucht wird Literatur zum Thema XY) vs. formal (Was ist in der neuesten Ausgabe der Zeitschrift XY erschienen?)
- Zielgruppe  
An welche Zielgruppe/Nutzergruppe richtet sich das Informationsangebot?  
z. B. WissenschaftlerInnen, Studierende, Lehrende, Lernende, interessierte Öffentlichkeit
- Kommunikationsart  
Um welche Form der Kommunikation handelt es sich?  
z. B. formelle wissenschaftliche Kommunikation, informelle wissenschaftliche Kommunikation; einseitige vs. zweiseitige Kommunikation
- Qualitätskontrolle  
Erfolgt eine Kontrolle der wissenschaftlichen Qualität des Angebots? Wenn ja, wie?  
z. B. ja/nein, Verfahren der Qualitätskontrolle (z. B. Peer Review oder Auswahl durch eine Redaktion)

---

<sup>26</sup> [http://intranet.vascoda.de/fileadmin/vascoda-](http://intranet.vascoda.de/fileadmin/vascoda-storage/Themen/Fachportale/Core_Set/Dokumente/core_set.pdf)

[storage/Themen/Fachportale/Core\\_Set/Dokumente/core\\_set.pdf](http://intranet.vascoda.de/fileadmin/vascoda-storage/Themen/Fachportale/Core_Set/Dokumente/core_set.pdf)

<sup>27</sup> Morville & Rosenfeld unterscheiden drei Arten von Rechercheverhalten („information-seeking behavior“): Suchen („searching“), Blättern („browsing“) und Anfragen („asking“). Suchen ist durch die Eingabe von Suchtermen charakterisiert, Blättern durch das Navigieren von Link zu Link und Anfragen durch das Formulieren von Fragen an Personen per E-Mail, Formular, Chat etc. (vgl. 2006, 42).

- Informationsart  
Welche Informationsart beinhaltet das Angebot?  
z. B. Metadaten, Volltexte, etc.
- Aktivitäten vs. Ergebnisse  
Beinhaltet das Informationsangebot Forschungs*aktivitäten* oder Forschungs*ergebnisse*?  
Dazu zählen z. B. Projekte (Aktivitäten), Publikationen, Primärdaten (Ergebnisse).
- Struktur  
Bietet das Angebot Informationen zur Struktur des Faches?  
Hier geht es um die Akteure innerhalb eines Faches, also z. B. Personen, Institutionen, Netzwerke, aber auch um Zeitschriftenprofile, die Auskunft darüber geben, wo innerhalb eines Faches publiziert wird.

Anhand dieser Fragen erfolgte eine grundsätzliche Einteilung der Module in vier Kategorien:

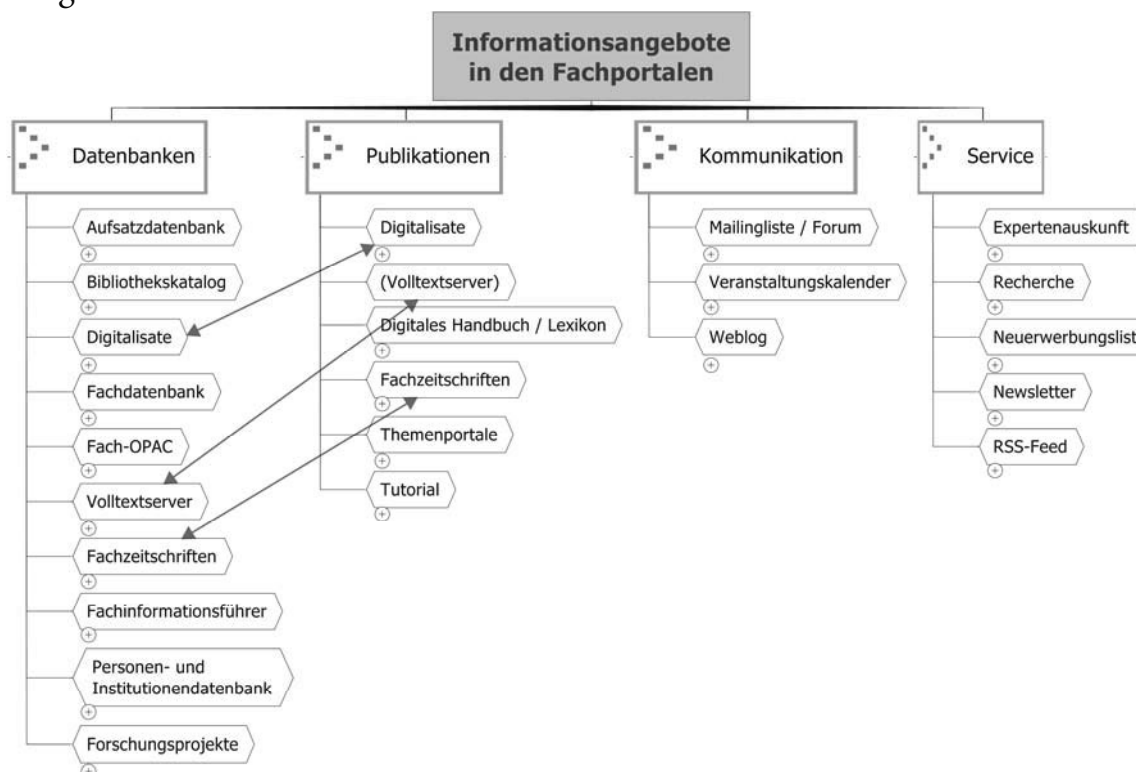


Abbildung 5: Gruppierung der vorhandenen Informationsangebote in den Fachportalen

Die folgende Tabelle gibt einen Überblick darüber, welche Kriterien in welcher Ausprägung den einzelnen Gruppen zugeordnet werden können:



	<i>Datenbanken</i>	<i>Publikationen</i>	<i>Kommunikation</i>	<i>Service</i>
primäre Interaktion	Suchen	Blättern	Feed-back/Austausch	Anfragen
Informationsbedürfnis	thematisch	formal	thematisch oder formal	thematisch oder formal
Zielgruppe(n)	alle	alle	alle	alle
Kommunikationsart	formell wiss.	formell wiss.	informell wiss.	informell wiss.
Qualitätskontrolle	ja	ja	nein	ja
Informationsart	Metadaten und Volltexte	Inhaltsverzeichnisse und Texte	Thema, Zeit, [Ort]	Anfrage/Angebot
Aktivitäten vs. Ergebnisse	Forschungsaktivitäten und –ergebnisse	Forschungsergebnisse	#	#
Struktur	Informationen zur Struktur des Faches	#	#	#

Tabelle 4: Kriterien zur Gruppierung der Informationsangebote in den Fachportalen

Anzumerken ist, dass die Kriterien nicht immer trennscharf sind und auch nicht auf alle Module anwendbar, aber eine erste Orientierung und Hilfestellung zur Kategorisierung der Informationsangebote bieten. Dabei kann ein Modul durchaus mehreren Kategorien zugeordnet werden.

Die Gruppe **Datenbanken** beinhaltet den großen Block der Literaturinformation. Dabei handelt es sich um qualitätsgeprüfte publizierte Fachinformation in Form von Literaturnachweisen und digitalen Volltexten. Zu finden ist dieser Informationstyp in den Modulen Aufsatzdatenbank, Bibliothekskatalog, Digitalisate, Fachdatenbank, Fach-OPAC, Fachzeitschriften, Volltextserver und teilweise auch im Modul Fachinformationsführer. Je nach Fach kommen auch noch Datenbanken hinzu, die Primärinformation (z. B. Umfragedaten, Werkstoffdaten, Chemiedaten etc.) enthalten. Bisher ist diese Informationsart in den Fachportalen aber noch selten vertreten. Neben den Modulen, die Literaturinformation anbieten, wurden unter Datenbanken Module eingeordnet, die als Wegweiser durch die Struktur des jeweiligen Faches dienen sowie wissenschaftliche Aktivitäten und Ergebnisse erfassen: Dazu gehören die Fachinformationsführer, die Personen- und Institutionenda-

tenbanken sowie Forschungsprojekte. Der Fachinformationsführer nimmt insofern eine Sonderrolle ein, da einige Fachportale in diesem Modul nicht nur Metadaten zu institutionellen oder thematischen Websites erfassen, sondern auch zu elektronischen Volltexten. Die Informationsangebote, die die Module bieten, sind mit einer variierenden Anzahl von Metadaten unterschiedlich tief formal und inhaltlich erschlossen (vgl. Mayr 2006). Gemeinsam ist allen genannten Modulen, dass der Hauptzugang für den Nutzer in der Suche mit Termen (Deskriptorensuche) (vgl. Krause & Mayr 2006, 42) besteht. Beispiele für das Ergebnis der Suche in der Gruppe Datenbanken sind im Fall der Literaturinformation Volltexte bzw. Verweise auf Volltexte (plus Verfügbarkeitsinformationen) und im Fall der Forschungsinformation mit Metadaten angereicherte Verweise auf institutionelle Websites.

In der Gruppe **Publikationen** sind elektronische Veröffentlichungen eingeordnet, die im Fachportal erscheinen, also redaktionelle Inhalte im Volltext. Zurzeit ist dieser Bereich in den Fachportalen (auch aus Gründen der Ressourcenknappheit) noch wenig vertreten. Es existieren jedoch bereits verschiedene Ansätze, die – auch im Sinne der DFG<sup>28</sup> – weiter ausgebaut werden sollten. Als Beispiele können die Publikation des Rezensionssjournals KUNSTFORM im Fachportal arthistoricum.net<sup>29</sup> oder die Rubrik „Themen“ im Fachportal historicum.net<sup>30</sup> genannt werden. Möglich ist das Angebot von Open-Access-Fachzeitschriften, Rezensionen, Lexikaartikeln, Themenportalen oder E-Learning-Materialien in Kooperation mit der jeweiligen Fachcommunity. Der Zugriff auf die Publikationen erfolgt im Unterschied zu den Datenbanken vorrangig über das Browsing von z. B. Inhaltsverzeichnissen (was eine Suche jedoch nicht ausschließt). Auf unterster Ebene des Browsing wird dem Nutzer dann ein Volltext angeboten.

In die Gruppe **Kommunikation** fallen Module wie Mailinglisten und Foren, Veranstaltungskalender und Weblogs. Sie zeichnen sich dadurch aus, dass sie dem Nutzer die Möglichkeit bieten, sich aktiv einzubringen. Strukturiert sind die Inhalte dieser Module durch die Angabe eines Themas/Titels, Autors/Veranstalters sowie ggf. die Angabe von Zeit und Ort. Es können Verweise auf andere Inhalte enthalten sein, und dem Nutzer kann die Möglichkeit zum Kommentieren und Bewerten geboten werden. Zukünftig sind in dieser Gruppe weitere Web 2.0-Angebote wie Podcasts oder Video-Blogs (Vlogs) etc. denkbar.

---

<sup>28</sup> Vgl. DFG-Positionspapier zu den Förderschwerpunkten wissenschaftlicher Literaturversorgungs- und Informationssysteme (2006, 5).

<sup>29</sup> [http://www.arthistoricum.net/epublishing/kunstform/?no\\_cache=1](http://www.arthistoricum.net/epublishing/kunstform/?no_cache=1)

<sup>30</sup> <http://www.historicum.net/themen>

In der Gruppe **Service** finden sich Module wie Expertenauskunft und Recherche sowie Neuerwerbungslisten, Newsletter und RSS-Feed. Die Portale bieten hier zusätzliche Dienstleistungen zur Informationsvermittlung an, die vom Nutzer explizit angefordert werden müssen und die er dann einmalig oder regelmäßig erhält. Im Falle der beiden Module Expertenauskunft und Recherche handelt es sich um Angebote, die vom Nutzer eine an Personen gerichtete Anfrage erfordern.

Das Ziel ist die semantische Integration der Module in den unterschiedlichen Gruppen. Dies betrifft in erster Linie die beiden Gruppen Datenbanken und Publikationen: In den Publikationen sollte es z. B. möglich sein, durch Klick auf den Autornamen eines Zeitschriftenartikels alle Verweise aus den Datenbanken zu dieser Person zurückgeliefert zu bekommen.

Die Einteilung der Module in die genannten vier Gruppen bildet die Basis der Informationsarchitektur der Fachportale. Als nächstes muss die Frage beantwortet werden, wie diese strukturelle Gestaltung unter Berücksichtigung softwareergonomischer Kriterien in eine Oberfläche umgesetzt werden kann.

#### **4.3 Umsetzung der Informationsarchitektur auf softwareergonomischer Ebene**

Die besondere Herausforderung bei der Umsetzung der Informationsarchitektur in eine auf einem konsistenten Gestaltungs- und Bedienkonzept basierende Oberfläche besteht darin, die Komplexität<sup>31</sup> der Fachportale in einer Weise abzubilden, die den NutzerInnen einen effizienten Zugriff auf die Informationsangebote erlaubt. Derzeit ist dies in den Fachportalen nicht der Fall. Aus der Analyse der geschilderten heterogenen Angebotsstruktur lässt sich der Schluss ziehen, dass die mangelnde Übersichtlichkeit und Komplexität der Fachportale in erster Linie durch die Vermischung von Website-Navigation und inhaltlicher Navigation entstehen. Die Website-Navigation bietet eine feste Sicht auf die Website und unterliegt softwareergonomischen Richtlinien z. B. hinsichtlich Menütiefe und -breite<sup>32</sup>. Gängige softwareergonomische Richtlinien empfehlen eine flache Menüstruktur, d.h. maximale Dreistufigkeit, die sich jedoch mit der komplexen Struktur des In-

---

<sup>31</sup> Gonçalves et al. (2004) und Candela et al. (2007) identifizieren extreme Komplexität als hervorstechendsten Merkmal digitaler Bibliotheken.

<sup>32</sup> Zur Frage, wie viele Ebenen (Tiefe) und wie viele Einträge auf den einzelnen Ebenen (Breite) ein Menü optimaler Weise haben sollte, gibt es eine Vielzahl von Studien: vgl. z. B. Jacko & Salvendi (1996), Kiger (1984) und Miller (1981) zur Gestaltung von Menüs in grafischen Oberflächen allgemein sowie Larson & Czerwinski (1998) und Zaphiris & Mtei (1997) zur Menügestaltung speziell im Webdesign.

formationsangebots eines Fachportals nicht vereinbaren lässt. Die inhaltliche Navigation bietet dagegen keine softwareergonomische Sicht auf die Welt, sondern eine dokumentarische/informationsarchitektonische Sicht. Sie ist wesentlich komplexer und umfasst u. U. eine tiefe Hierarchie. Aber auch diese Hierarchie bildet nur eine Sicht auf die Welt ab, die nie allen Informationsbedürfnissen der Nutzer gerecht werden kann. Für diese Problematik existieren zwar Lösungen (vgl. z. B. die Facettenklassifikation im Statistikinformationssystem ELVIRA, beschrieben in Stempfhuber 2003, 109ff.), die aber mit Webtechnologien schwierig umzusetzen sind. Deshalb schlagen wir die Trennung von Website- und Inhaltsnavigation vor. Während die Website-Navigation eine statische Sicht bietet, soll die inhaltliche Navigation dynamische Sichten ermöglichen, die vom Nutzer je nach Informationsbedürfnis gewechselt werden können. Auf diese Weise entsteht ein eigenständiges Zugangssystem für die dokumentarischen Inhalte des Fachportals.

Als softwareergonomisches Gestaltungsmittel für die Umsetzung der dynamischen Sichten auf die Inhalte des Fachportals wurde das Prinzip des Produktkatalogs gewählt. Wirth (2004) nennt den Produktkatalog als gängiges Beispiel für die Umsetzung von Hierarchien im Web (vgl. 344). Er wird vor allem im Bereich E-Commerce (z. B. Online-Shops) eingesetzt, um den NutzerInnen einen effizienten Zugriff auf vielfältige Angebote zu ermöglichen.

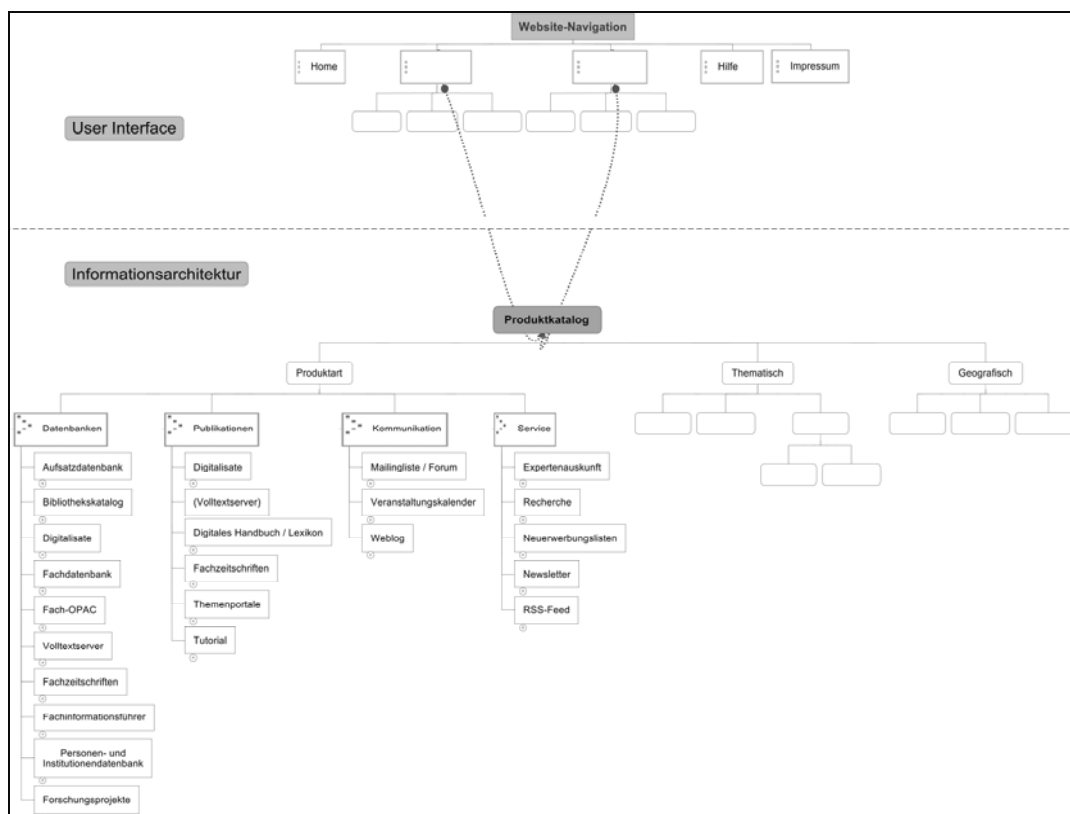


Abbildung 6: Integrierte Informationsarchitektur der Fachportale

Aus den vorangegangenen Überlegungen ergibt sich das in Abbildung 6 gezeigte Modell.

Deutlich sichtbar ist die Trennung von Website-Navigation (oben) und Produktkatalog (unten). Dabei gilt es, folgende Fragen zu beantworten: Wo hört die Website-Navigation auf? Wo beginnt der Produktkatalog? Was sind Menüpunkte im Sinne der Site-Navigation, was im Sinne der Informationsarchitektur? Im linken Zweig des Produktkatalogs findet sich die an den Produktarten orientierte Sicht auf die Informationsangebote aus Kap. 4.2 wieder. Mögliche weitere Sichten wären eine thematische und eine geografische.

Das Modell ist der Versuch, ein Verfahren zur Vereinheitlichung der Fachportale zu entwickeln, dass die Komplexität reduziert, unterschiedliche Sichtweisen unterstützt und ein einheitliches mentales Modell impliziert, das einerseits übergreifende Gemeinsamkeiten hervorhebt, aber andererseits den Fachportalen individuellen Gestaltungsspielraum erlaubt.

## **5      Ausblick**

Um die Benutzerfreundlichkeit des vorgeschlagenen Referenzmodells evaluieren zu können, wird zurzeit ein Prototyp entwickelt, der die Trennung von Website-Navigation und inhaltlicher Navigation in Form eines Produktkatalogs beispielhaft umsetzt. Dieser Prototyp wird durch einen Benutzertest evaluiert. Auf Basis der Testergebnisse wird ein Musterportal sowie ein Corporate Design Styleguide entwickelt. Beides kann von den in vascoda integrierten Fachportalen genutzt und in definierten Grenzen an das konkrete Portal angepasst werden, um sich auf ergonomischer und ästhetischer Ebene anzunähern.

## **6      Literatur**

- Arndt, Henrik (2006): Integrierte Informationsarchitektur: Die erfolgreiche Konzeption professioneller Websites: Springer.
- Balzert, Helmut et al. (Hrsg.) (1988): Einführung in die Software-Ergonomie. Berlin [u a.]: de Gruyter. (Mensch-Computer-Kommunikation Grundwissen; 1)
- Candela, Leonardo; Castelli, Donatella; Pagano, Pasquale; Thanos, Constantino; Ioannidis, Yannis; Koutrika, Georgia; Ross, Seamus; Schek, Hans-Jörg; Schuldt, Heiko (2007): Setting the Foundations of Digital Libraries: The DELOS Manifesto. In: D-Lib Magazine 13, Nr. 3-4.  
URL: <http://www.dlib.org/dlib/march07/castelli/03castelli.html>

- Deutsche Forschungsgemeinschaft (2003): Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung. Teil I: Informationsverhalten und Informationsbedarf der Wissenschaft (Juni 2003). Bearb. von Boekhorst, Peter te; Kayß, Matthias; Poll, Roswitha.  
URL: [http://www.dfg.de/forschungsfoerderung/wissenschaftliche\\_infrastruktur/lis/download/ssg\\_bericht\\_teil\\_1.pdf](http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/ssg_bericht_teil_1.pdf)
- Deutsche Forschungsgemeinschaft (2006): DFG-Positionspapier: Wissenschaftliche Literaturversorgungs- und Informationssysteme: Schwerpunkte der Förderung bis 2015.  
URL: [http://www.dfg.de/forschungsfoerderung/wissenschaftliche\\_infrastruktur/lis/download/positionspapier.pdf](http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/positionspapier.pdf)
- Fast, Karl (2006): IA Column: The Confluence of Research and Practice in Information Architecture. In: ASIS&T Bulletin, June/July 2006.  
URL: <http://www.asis.org/Bulletin/Jun-06/fast.html>
- Germonprez, Matt; Zigurs, Ilze (2003): Causal Factors for Web Site Complexity. In: Sprouts: Working Papers on Information Environments, Systems and Organizations 3, Nr. 2, S. 107-121.  
URL: <http://sprouts.case.edu/2003/030205.pdf>
- Gonçalves, Marcos André; Fox, Edward A.; Watson, Layne T.; Kipp, Neill A. (2004): Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries. In: ACM Transactions on Information Systems (TOIS) 22, Nr. 2, S. 270-312
- Heinz, Sabine; Pianos, Tamara (2007): vascoda – ein Netzwerk von Fachportalen. S. xxx-xxx. [erscheint] In: Stempfhuber, Maximilian (Hrsg.): Lokal – Global: Vernetzung wissenschaftlicher Infrastrukturen – 12. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaften in Deutschland. Bonn: IZ Sozialwissenschaften.
- Herczeg, Michael (1994): Software-Ergonomie: Grundlagen der Mensch-Computer-Kommunikation. Bonn [u a.]: Addison-Wesley.
- Jacko, J. A.; Salvendi G. (1996): Hierarchical menu design: breadth, depth, and task complexity. In: Perceptual and Motor Skills, Nr. 82, S. 1187-1201.
- JASIST (2002): Journal of the American Society for Information Science and Technology (JASIST) 53, Nr. 10.  
URL: <http://www.asis.org/Publications/JASIS/vol53n10.html>
- Kiger, J. (1984): The depth/breadth tradeoff in the design of menu-driven user interfaces. In: International Journal of Man-Machine Studies. Bd. 20, Nr. 2, S. 201-214.
- Krause, Jürgen (2004): Software-Ergonomie. In: Kuhlen, Rainer; Seeger, Thomas; Strauch, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis; Bd. 1. München: Saur. S. 371-375.
- Krause, Jürgen; Mayr, Philipp (2006): Allgemeiner Bibliothekszugang und Varianten der Suchtypologie – Konsequenzen für die Modellbildung in vascoda. Bonn: Informationszentrum Sozialwissenschaften. (IZ-Arbeitsbericht Nr. 38)  
URL: [http://www.gesis.org/Publikationen/Berichte/IZ\\_Arbeitsberichte/pdf/ab\\_38.pdf](http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_38.pdf)

- Larson, K.; Czerwinski, M. (1998): Web page design: implications of memory, structure and scent for information retrieval. In: Proceedings of CHI, ACM, New York. S. 25-32.
- Lynch, Patrick J.; Horton, Sarah (2002): Web style guide: basic design principles for creating web sites. Yale University Press.  
URL: <http://www.webstyleguide.com>, Stand: 12.07.2005
- MacKenzie, M.; Laskey, K.; McCabe, F.; Brown, P.; Metz, R. (2006): Reference Model for Service Oriented Architecture. OASIS Committee Draft 1.0.  
URL:  
<http://www.oasis-open.org/committees/download.php/16587/wd-soa-rm-cd1ED.pdf>
- Mayr, Philipp (2006): Informationsangebote für das Wissenschaftsportal vascoda – eine Bestandsaufnahme. Bonn: IZ Sozialwissenschaften. (IZ-Arbeitsbericht; Nr. 37)  
URL: [http://www.gesis.org/Publikationen/Berichte/IZ\\_Arbeitsberichte/pdf/ab\\_37.pdf](http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_37.pdf)
- Miller, D. P. (1981): The depth/breadth tradeoff in hierarchical computer menus. In: Proceedings of the Human Factors Society. S. 296-300.
- Morville, Peter; Rosenfeld, Louis (2006): Information Architecture for the World Wide Web: Designing Large-Scale Web Sites. 3. Aufl. Beijing [u a.]: O'Reilly.
- Pianos, Tamara; Trauzold, Silke (2006): Ergebnisse der Befragung unter den Fachportalen. [internes Papier, im vascoda-Intranet für Vereinsmitglieder zugänglich]
- Rosenfeld, Louis; Morville, Peter (1998): Information architecture for the World Wide Web: designing large-scale Web sites. Beijing [u a.]: O'Reilly.
- Shneiderman, Ben (1998): Designing the user interface: strategies for effective human-computer interaction. Reading, Mass. [u a.]: Addison-Wesley.
- Stempfhuber, Maximilian (2003): Objektorientierte Dynamische Benutzungsoberflächen – ODIN: Behandlung semantischer und struktureller Heterogenität in Informationssystemen mit den Mitteln der Softwareergonomie. Bonn: IZ Sozialwissenschaften. (Forschungsberichte; 6)
- Wirth, Thomas (2004): Missing Links. Über gutes Webdesign. München: Hanser.
- Wurman, Richard Saul; Bradford, Peter (1996): Information architects. Zürich: Graphis Press.
- Zaphiris, P. & Mtei, L. (1997): Depth vs breadth in the arrangement of web links.  
URL: <http://otal.umd.edu/SHORE/bso4>
- ZB MED (Hrsg.) (2006): Usability-Studie zum Internetportal vascoda. November 2006.  
URL: [http://www.dl-forum.de/dateien/vascoda\\_Usability\\_Studie\\_2006-Nov.pdf](http://www.dl-forum.de/dateien/vascoda_Usability_Studie_2006-Nov.pdf)





# Heterogenität in wissenschaftlichen Fachdatenportalen\*

*Stefan Baerisch*

Informationszentrum Sozialwissenschaften  
Lennéstr. 30, 53113 Bonn  
*bs@iz-soz.de*

## **Zusammenfassung**

Bei der Bereitstellung von Informationsbeständen verschiedener Anbieter in einem gemeinsamen Portal müssen sich Anbieter verschiedenen Anforderungen bei der Behandlung von Heterogenität stellen. Neben der Zusammenführung und Vereinheitlichung auf struktureller Ebene sind auch Fragestellungen der semantischen Heterogenität zu beachten. Eine weitere Heterogenitätsdimension liegt in den unterschiedlichen Nutzergruppen, die auf ein solches Portal zugreifen. In diesem Papier gehen wir auf Konzepte zur Behandlung verschiedener Heterogenitätsdimensionen ein und diskutieren, wie diese Konzepte in einem Portal angewandt werden können. Schwerpunkte sind die Behandlung von struktureller Heterogenität durch semistrukturierte Datenformate und inkrementelle Integrationsprozesse sowie die Behandlung der semantischen Heterogenität durch Crosskonkordanzen. Abschließend wird die Umsetzung dieser Konzepte am Beispiel des sozialwissenschaftlichen Fachportals SOWIPOINT erläutert.

## **1 Einleitung**

Eines der zentralen Anliegen der aktuellen Informationslandschaft ist die Integration vorhandener Informationen zu einem gemeinsamen Zugang. Für einen Nutzer von Informationsdiensten ist es nicht zu leisten, für die Recherche zu einem Thema in mehreren Schritten zuerst potentielle Informationsquellen zu identifizieren, diese zu evaluieren und dann erst anzufragen. Diese Problemstellung wird in der Praxis noch verschärft, da die angebotenen Informationsbestände Aktualisierungen unterliegen, somit also eine Wiederholung von Anfragen in regelmäßigen Abständen notwendig wäre. Bei Betrachtung der Anzahl von im Deep Web (siehe [Ragha-

---

\* Veröffentlicht in: OSSWALD, Achim; STEMPFHUBER, Maximilian; WOLFF, Christian (Hrsg.) (2007). Open Innovation. Proc. 13. Jahrestagung der IuK-Initiative Wissenschaft. Konstanz: UVK, 509-518.

vano 1]) verfügbaren Quellen wird offenbar, dass ein manuelles Anfragen von Quellen nicht wünschenswert sein kann.

Eine erste Herausforderung bei der Behandlung mehrerer Quellen ist somit die gemeinsame Anfrage dieser Quellen. Zur integrierten Anfrage verteilter Datenquellen existieren in Forschung und Praxis eine Reihe von Ansätzen, etwa die Verwendung von Metasuchen und Föderierter Suche (siehe [Baeza-Yates99]) und verteilten Datenbanken. Eine alternative Möglichkeit besteht in der zentralen Indizierung der zugänglich zu machenden Informationen. Ist eine gemeinsame Anfrage auf allen Informationsbeständen etabliert, stellt sich als nächste Herausforderung die Integration von strukturellen Unterschieden im Datenbestand: Unterschiedliche Datenanbieter verwenden abweichende Modellierungen der beschriebenen Entitäten, insbesondere von Bedeutung ist die Verfügbarkeit von Daten in verschiedenen Detailgraden. Semantische Heterogenität tritt auf, wenn sich die Konzepte und Vokabulare unterscheiden, die zur Beschreibung von Themen verwandt werden. Schlagwortlisten und Thesauri mit verschiedenem Umfang und unterschiedlichen Schwerpunktsetzungen müssen einem Anwender zur Verfügung gestellt werden, ohne dass dieser Kenntnis über die Details der jeweiligen Vokabulare hat.

Ein letzter Punkt der Heterogenität betrifft nicht die bereitgestellten Informationen als solche, sondern die Art und Weise, wie verschiedene Anwendergruppen ihr jeweiliges Informationsbedürfnis an den integrierten Datenbestand ausdrücken. Eine Übersicht der verschiedenen Heterogenitätsarten bietet die Abbildung 1.

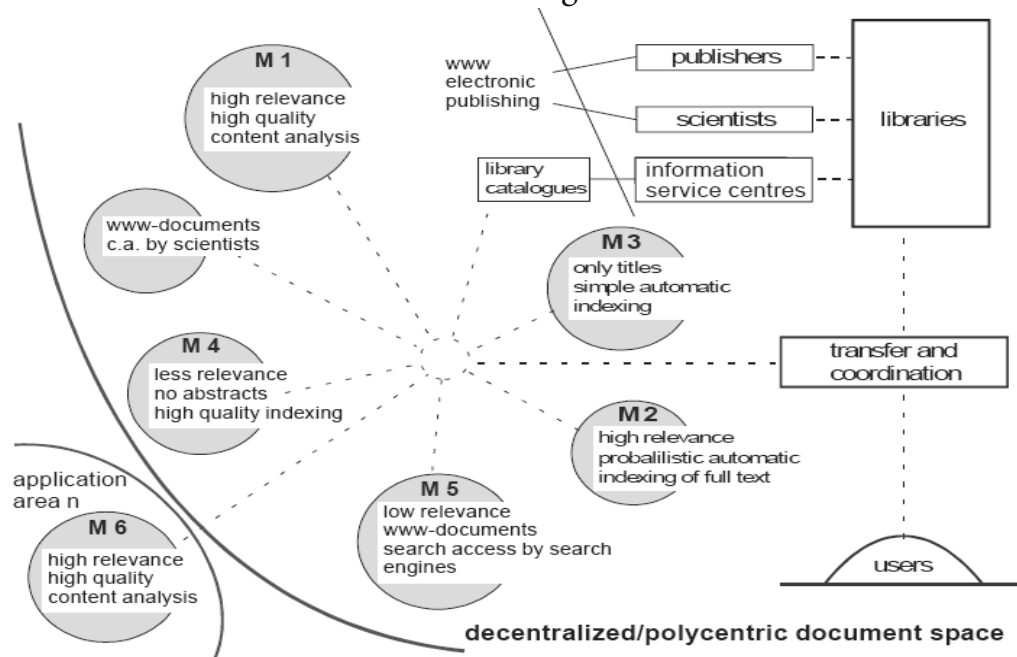


Abbildung 1: Strukturell, semantisch und qualitativ heterogene Datenbestände nach [Krause2004]

In diesem Papier gehen wir auf diese verschiedenen Aspekte oder Dimensionen der Heterogenität ein, die in einem Fachportal behandelt werden müssen. Als Grundlage diskutieren wir die strukturelle Heterogenität in Abschnitt 2 und stellen Ansätze und Integrationsverfahren vor. In diesem Kontext gehen wir auch auf grundsätzliche Aspekte bei der Zusammenführung von Daten ein. Abschnitt 3 geht auf die Behandlung der semantischen Heterogenität ein, im Kern der Betrachtung stehen intellektuelle und statistische Verfahren zur Behandlung zur Erstellung von Crosskonkordanzen. Abschnitt 4 diskutiert den Umgang mit heterogenen Nutzergruppen durch die Mittel der Softwareergonomie und der Oberflächengestaltung. Abschnitt 5 stellt abschließend die gemeinsame Anwendung der diskutierten Konzepte vor, den Hintergrund der Darstellung bildet das wissenschaftliche Fachportal SOWIPORT; vor diesem Hintergrund stellen wir auch unsere Betrachtung von der Informationsintegration als Prozess vor.

## **2 Datenzusammenführung und Strukturelle Integration**

Der Wunsch, verteilte Datenbestände anzufragen, ist seit langem Triebfeder und Thema der Forschung. Unterschieden werden muss hier zwischen der Informationsintegration und der Datenintegration. Die Datenintegration betrachtet vorrangig die Zusammenführung von verteilten, unterschiedlichen Datenbanken, wobei hier im Gegensatz zur Datenbankföderation der lesende Zugriff im Mittelpunkt steht. Ein häufiges Szenario für die Anwendung von Datenintegrationsverfahren im wirtschaftlichen Umfeld ist die Zusammenführung verschiedener Datenbestände in einem Datawarehouse. Im Rahmen der Informationsintegration stehen im Vergleich zur Datenintegration eher semistrukturierte Daten im Mittelpunkt der Betrachtung, da die Daten sind weniger für die automatische Verarbeitung als für die menschliche Information gedacht sind. Eine Folge ist, die Verwendung von automatischen Verfahren zum Schema-Matching erschwert wird. Auch die Anwendung von auf Dateninstanzen basierenden Verfahren wird durch die breite Spannweite an Ansetzungsformen erschwert.

Die genannten Eigenschaften der bei der Informationsintegration betrachteten Daten erzwingen in der Praxis einen intellektuellen Ansatz zur Datenintegration, was die Frage nach geeigneten Verfahren und Ansätzen aufwirft. Ein aus der Datenintegration entnommener, auch für die strukturelle Informationsintegration anwendbarer Ansatz ist das Konzept der 'Global as View' (siehe[Halevy2006]). Hierbei werden die zu integrierenden Datenbestände in einem gemeinsamen Integrationschema zusammengefasst. Wrapper nehmen für jeden Datenbestand die Überfüh-

rung in ein gemeinsames Schema vor, führen also entweder Daten zusammen, wenn das Integrationsschema weniger feingranular ist als das Quellschema oder nehmen die Aufspaltung oder Aufbereitung von weniger fein strukturierten Daten vor.

Die Zusammenführung von Daten und der Einsatz von Wrappern kann sowohl im Rahmen einer einmaligen Datenaufbereitung geschehen als auch zum Anfragezeitpunkt. Werden Aspekte wie das Caching vernachlässigt, bedeutet die letztere Lösung eine Erhöhung der Anfragelast auf das jeweilige Einzelsystem. Insbesondere wenn Verfahren zur Relevanzermittlung im Sinne eines statistischen Information Retrieval zum Einsatz kommen, kann der Ressourcenverbrauch einer Anfrage der Integrationsplattform an die Einzelsysteme jenen aus regulären Nutzeranfragen bei Weitem übersteigen. Bei einer solchen Anfrage ist in der Regel neben den eigentlich nutzerrelevanten Informationen die Übertragung von statistischen Hilfsinformationen notwendig. Die Zusammenführung aller Datenbestände an einer Stelle hat unter den Aspekten der Anfrageperformanz Vorteile, ein potentiell Problem ist jedoch neben organisatorischen Aspekten die Aktualität des Datenbestands, da die Konvertierung von Daten zum Indizierungszeitpunkt stattfindet.

Generell liegt die Herausforderung des Prozesses der strukturellen Informationsintegration in der Definition eines geeigneten Integrationsschemas. Die Eignung ist hier sowohl von den Anforderungen abhängig, die sich an das integrierte Datenformat stellen als auch von den zu integrierenden Daten.

### **3 Semantische Heterogenität**

Der Begriff der semantischen Heterogenität bezeichnet unter anderem die Verwendung von verschiedenen Verschlagwortungsvokabularen in den Informationsbeständen eines integrierten Informationssystems. Zur Behandlung der semantischen Heterogenität existieren verschiedene Ansätze. Zu unterscheiden sind in erster Linie die intellektuelle Erstellung von Crosskonkordanzen und die Verwendung von statistischen Verfahren zur Zuordnung von Termen zwischen Vokabularen. Mayr und Walter stellen verschiedene Verfahren vor und gehen auf Methoden zur Evaluation ein (siehe [Mayr07]). Durch die Verwendung von Crosskonkordanzen wird ein Graph definiert, der entlang seiner Kanten die Umsetzung von Anfragen in fremde Verschlagwortungsvokabulare erlaubt, Abbildung 2 stellt einen solchen Graphen dar. Abbildung 3 zeigt die Umwandlung einer Anfrage nach "Bildungseinrichtung", eines Terms auf dem Thesaurus Sozialwissenschaften, auf verschiedene Zielvokabulare, hier die Schlagwortnormdatei und den Psynindex-Thesaurus.

Im Rahmen der semantischen Integration muss auch die Behandlung von Datenbeständen unterschiedlicher Qualität betrachtet werden, insbesondere bezüglich der Qualität der inhaltlichen Erschließung, Krause definiert hierfür das Schalenmodell (siehe [Krause2004])

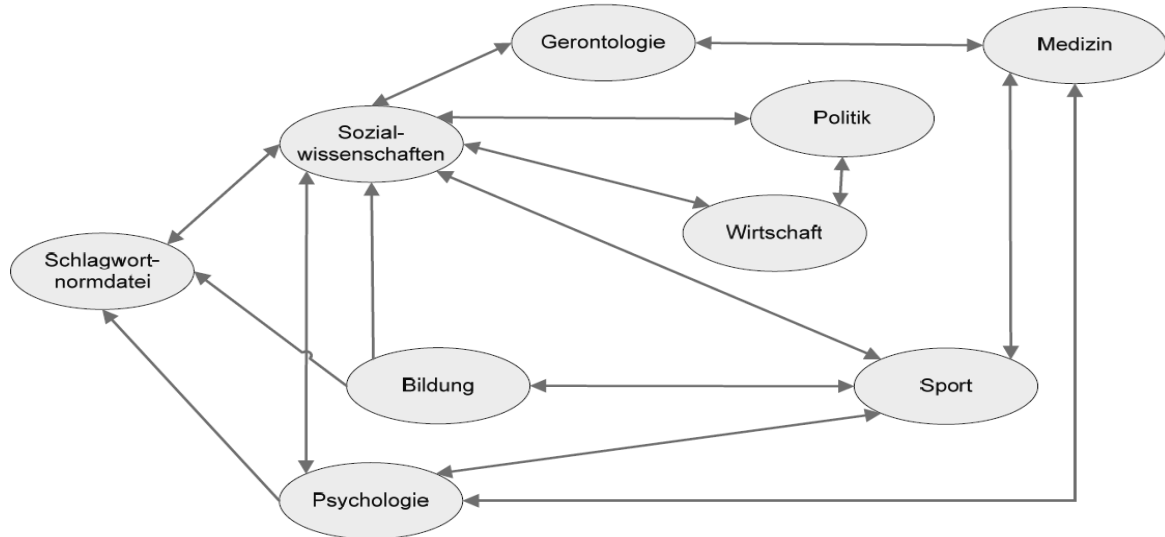


Abbildung 2: Von Crosskondordanzen aufgespannter Graph zur Behandlung semantischer Heterogenität aus [Mayr2007]

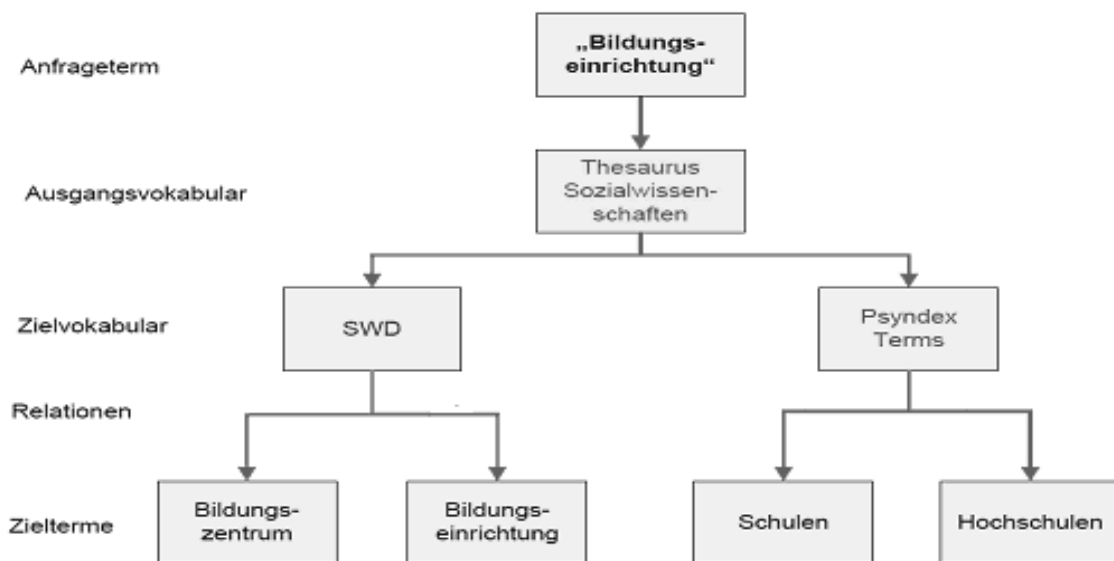


Abbildung 3: Beispielanfrage bei Verwendung von Crosskonkordanzen nach [Mayr2007, modifiziert]

## 4 Lösungsansätze für verschiedene Anwendergruppen

Die Anwendergruppen, die ein System nutzen, sind in hohem Maß unterschiedlich. Geht man davon aus, dass das betrachtete Integrationsprojekt vertikal orientiert ist,

also inhaltlich oder thematisch spezialisiert ist, so bestehen noch immer Unterschiede im konkreten Informationsbedürfnis und in den bekannten oder bevorzugten Methoden, dieses Informationsbedürfnis auszudrücken. Internetsuchmaschinen wie Google haben bezüglich der Anfragedarstellung einen Standard gesetzt, an dem sich Informationsanbieter wohl oder übel messen lassen müssen, zugleich ist die Verwendung einer einfachen Suchzeile, wie sie von Google propagiert wird, nicht in allen Fällen geeignet für die Formulierung komplexer Anfragen.

Ein Mittel, die verschiedenen Nutzererwartungen und -anforderungen zu behandeln, ist die Bereitstellung von spezifischen Anfrageschnittstellen für unterschiedliche Nutzergruppen. Typische Beispiele sind hier die *Erweiterte Suche* oder die *Expertensuche*, die von vielen Informationsanbietern in Ergänzung zu einer *Einfachen Suche*, meist einer einzelnen Suchzeile, angeboten werden.

Eine Herausforderung an die verschiedenen Recherchezugänge ist die Beantwortung der Frage, wie ein Anwender bei der Formulierung seines Informationsbedürfnisses unterstützt werden kann. Ein Beispiel für eine solche Unterstützung ist die automatische Anfrageerweiterung, etwa die Umwandlung eingehender Nutzeranfragen mittels Crosskonkordanzen so dass der Recall einer Anfrage erhöht wird. Bei der Anwendung von Anfrageerweiterungen muss beachtet werden, dass diese für den Anwender nachvollziehbar sind und ihn gegebenenfalls in die Lage zu versetzen, in Zukunft selbst optimierte Anfragen zu erstellen.

Neben der Anfrageerweiterung besteht ein weiterer Ansatz, mit unterschiedlichen Voraussetzungen seitens der Anwender umzugehen in der Bereitstellung von Informationen über den Inhalt der Datenbank. Ein Mittel, dies zu leisten sind Facetten oder Navigatoren. Navigatoren bieten eine intuitive Möglichkeit, eine weit gefasste Anfrage einzuschränken; zugleich bieten sie Information zum Inhalt der Datenbank, die Nutzern bei der Formulierung von neuen Anfragen hilfreich sein können. Abbildung 4 zeigt drei Navigatoren aus dem SOWIPOINT Portal die es erlauben eine Treffermenge anhand der Quelldatenbank oder anhand von Personen oder Schlagworten aus den bei einer Anfrage gefundenen Dokumenten einzuschränken.

Eine Herausforderung, die nicht durch Navigatoren behandelt werden kann, ist die Darstellung der verbliebenen Heterogenität von Datenbeständen in Fällen, in denen die strukturelle oder semantische Heterogenität nicht in vollem Umfang behandelt werden konnte. So soll einem Anwender etwa konkret vermittelt werden, dass in einer bereitgestellten Literaturdatenbank grundsätzlich keine Zeitschriftenartikel enthalten sind oder welche Selektivität die konkreten Terme einer booleschen Suchanfrage aufweisen. Die Integration entsprechender Informationen in

Benutzungsoberflächen in einer für den Anwender verständlichen Form ist noch immer Gegenstand der Forschung, ein Beispiel für einen entsprechenden Ansatz ist ODIN (siehe [Stempfhuber2002]), dargestellt in Abbildung 5.



Abbildung 4: Navigatoren zur Einschränkung einer Treffermenge anhand von Personennamen, Schlagworten und der Quelldatenbank eines Ergebnisses

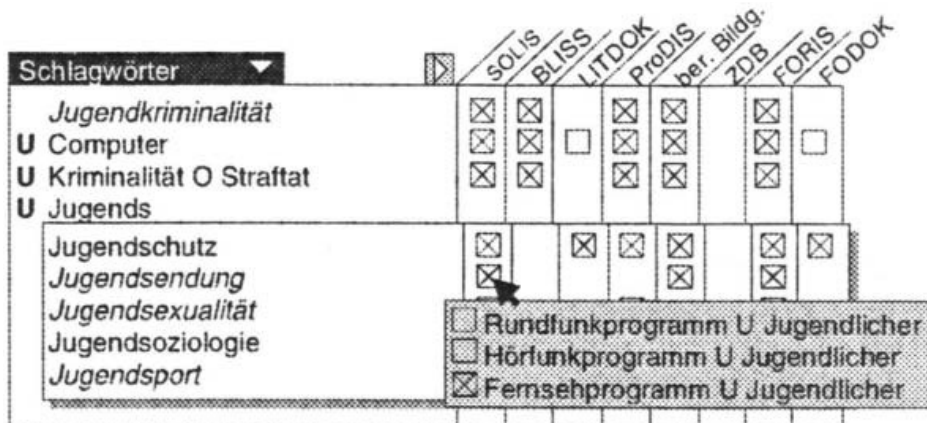


Abbildung 5: Behandlung semantischer Heterogenität durch die ODIN-Benutzungsoberfläche nach [Stempfhuber2002]

Eine Kernfrage moderner Informationssysteme ist die Frage, wie die Relevanz von Informationen in Bezug auf das Informationsbedürfnis eines Anwenders ermittelt werden kann. Gegenüber dem Booleschen Modell finden aktuell Modelle wie das statistische Information Retrieval oder das erweiterte boolesche Modell zunehmend Verbreitung. Diese Modelle bieten eine gestufte Bewertung der Relevanz von Ergebnisdokumenten, womit sich die Frage stellt, anhand welcher Eigenschaften eines

Dokuments sich die Relevanz ermitteln lässt. Das Information Retrieval sieht die Ermittlung eines Relevanzwertes anhand des Inhalts eines Dokuments vor, wobei in der Regel unterschiedliche Dokumentenbestandteile als unterschiedlich wichtig angesehen werden. So wird ein Term im Titel eines Dokuments als wichtiger erachtet als ein Treffer im Textkörper. Ergänzend zum Inhalt können die Positionen von Wörtern im Verhältnis zueinander betrachtet werden, so dass im Dokument nahe beieinander stehende Suchbegriffe eine höhere Relevanz erhalten. Topologische Verfahren wie Googles Pagerank betrachten die Verlinkung von Webseiten untereinander, im Bereich der Literaturrecherche können durch Zitationsanalysen ähnliche Verfahren angewandt werden. Externe Kriterien zur Dokumentenbewertung beziehen das Alter eines Dokuments in die Bewertung ein, wobei die Annahme zugrunde liegt, dass ältere Dokumente weniger relevant sind als jüngere, eine weitere Möglichkeit ist die Einbeziehung eines intellektuell oder automatisch vergebenen Qualitätswertes in die Relevanzermittlung.

Das Vorhandensein dieser Vielzahl von verschiedenen Relevanzkonzepten macht es schwer, ein allen Nutzeranforderungen entsprechendes, integriertes Konzept zu entwickeln. Vielmehr stellt sich die Frage, wie die einzelnen Aspekte der Relevanzermittlung einem Anwender gegenüber präsentiert werden können, um diesem die Auswahl einer geeigneten Methode der Relevanzermittlung zu gestatten.

## **5 Heterogenitätsbehandlung in SOWIPORT**

Bei der Konzeption und Umsetzung des sozialwissenschaftlichen Fachportals SOWIPORT sah sich das IZ Sozialwissenschaften mit den in diesem Papier genannten Dimensionen konfrontiert. Dieser Abschnitt stellt die Verfahren der Heterogenitätsbehandlung dar, die bei SOWIPORT Verwendung fanden und gibt einen Ausblick des geplanten weiteren Vorgehens.

SOWIPORT verfolgt einen zentralisierten Ansatz der Datenintegration: Alle im Portal durchsuchbaren Daten werden für eine Indizierung gesammelt und in ein gemeinsames Datenformat gebracht. Aufgrund der Heterogenität der enthaltenen Datenbestände und um bezüglich der zu unterstützenden Anfrageverfahren größtmögliche Flexibilität zu wahren, wurde ein Datenformat mit einem begrenzten Kernbestand von Konzepten formuliert das im Bedarfsfall für die Besonderheiten der jeweiligen Datenbestände spezialisiert werden kann. Die Zweistufigkeit von grundlegenden Konzepten zusammen mit der Möglichkeit der Spezialisierung erlaubt es, beim Entwurf des Schemas noch ungekannte Konzepte auszudrücken, hält das Datenformat zugleich aber übersichtlich.



Zurzeit wird für SOWIPORT eine zentrale Indizierungsstrategie verfolgt, alle beteiligten Partnerdatenbanken werden in einem zentralen, gemeinsamen Index zusammengefasst. Die Datenintegration findet dabei sowohl auf Ebene eines für die Darstellung verwandten, detaillierten XML-Datenformats als auch im Rahmen der Indizierung durch die FAST Suchmaschine statt, Abbildung 6 stellt den Datenfluss dar. Für die Zukunft ist hierüber hinaus auch die Einbindung externer Datenquellen vorgesehen die neben dem FAST Index angesprochen werden.

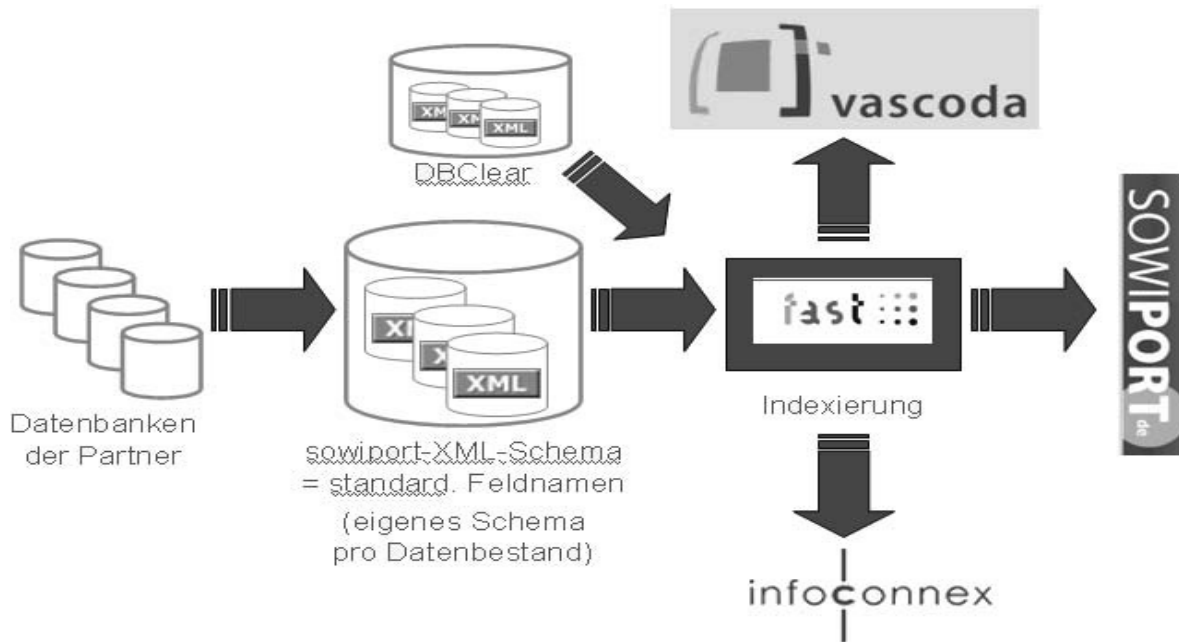


Abbildung 6: Verarbeitungs- und Indizierungspfade zur Bereitstellung heterogener Daten in SOWIPORT.

Bezüglich der Anfrage- und Ergebnisdarstellung stellt SOWIPORT zurzeit eine feldbasierte Suche in einer einfachen und einer erweiterten Version bereit um verschiedene Nutzergruppen anzusprechen. Mittels verschiedener, auswählbarer Kriterien zur Relevanzbestimmung werden diesbezüglich unterschiedliche Nutzeranforderungen unterstützt; die Verwendung von Facetten bietet die Möglichkeit zur Einschränkung von Suchanfragen.

Die Behandlung der semantischen Heterogenität geschieht mittels Crosskonkordanzen, diese werden bei der Anfrageabarbeitung automatisch auf Schlagwortsuchen angewandt. Die Verwendung der Crosskonkordanzen soll in weiteren Ausbaustufen durch die Bereitstellung eines Heterogenitätsservice erweitert werden, neben der Erweiterung von Anfragen wird dieser auch in der Lage sein, Anwender in einem interaktiven Prozess bei Anfragen mittels kontrollierten Vokabularen zu unterstützen.

## 6 Literatur

- R. Baeza-Yates; B. Ribeiro-Neto; et al.: Modern information retrieval, 1999 Verlag Addison-Wesley Harlow, England
- Alon Halevy; Anand Rajaraman; Joann Ordille: Data Integration: The Teenage Years. In: Conference on Very Large Data Bases '06, 12-15 September, 2006, Seoul, Korea.
- Krause, Jürgen: Standardization, Heterogeneity and the Quality of Content Analysis: a key conflict of digital libraries and its solution. In: IFLA Journal: Official Journal of the International Federation of Library Associations and Institutions 30, Nr. 4, 2004, S. 310 – 318
- Mayr, Philipp; Walter, Anne-Kathrin (erscheint): Zum Stand der Heterogenitätsbehandlung in vascoda: Bestandsaufnahme und Ausblick. In: Bibliothek & Information Deutschland (Hrsg.): 3. Leipziger Kongress für Information und Bibliothek, 19. – 22. März 2007. Leipzig: Verlag Dinges & Frick.
- S. Raghavan; H. Garcia-Molina: Crawling the Hidden Web. In: Proceedings of the 27<sup>th</sup> International Conference on Very Large Data Bases, 2001, Rom, S. 129 – 138
- Stempfhuber, Maximilian: Objektorientierte Dynamische Benutzeroberflächen ODIN, Dissertation an der Universität Koblenz-Landau, 2002